



**UNIVERSIDADE FEDERAL DA FRONTEIRA SUL  
CAMPUS DE CERRO LARGO  
CURSO DE ENGENHARIA AMBIENTAL E SANITÁRIA**

**ALEXANDRE LUIZ SCHÄFFER**

**ESTUDO DE INDICADORES DE AMBIENTE E SAÚDE UTILIZANDO A  
METODOLOGIA DE MODELOS MULTINÍVEIS NAS MICRORREGIÕES DO  
RIO GRANDE DO SUL**

**CERRO LARGO**

**2016**

**ALEXANDRE LUIZ SCHÄFFER**

**ESTUDO DE INDICADORES DE AMBIENTE E SAÚDE UTILIZANDO A  
METODOLOGIA DE MODELOS MULTINÍVEIS NAS MICRORREGIÕES DO  
RIO GRANDE DO SUL**

Trabalho de Conclusão de Curso apresentado ao  
Curso de Engenharia Ambiental e Sanitária da  
Universidade Federal da Fronteira Sul como  
requisito para obtenção do título de Engenheiro  
Ambiental e Sanitário.

**Orientadora:** Dra. Iara Denise Endruweit  
Battisti

**Coorientador:** Dr. Erikson Kaszubowski

**CERRO LARGO**

**2016**

## DGI/DGCI - Divisão de Gestão de Conhecimento e Inovação

Schäffer, Alexandre Luiz

Estudo de indicadores de ambiente e saúde utilizando a metodologia de modelos multiníveis nas microrregiões do Rio Grande do Sul/ Alexandre Luiz Schäffer. -- 2016. 94 f.

Orientador: Iara Denise Endruweit Battisti.

Co-orientador: Erikson Kaszubowski.

Trabalho de conclusão de curso (graduação) - Universidade Federal da Fronteira Sul, Curso de Engenharia Ambiental e Sanitária , Cerro Largo, RS, 2016.

1. Regressão multinível. 2. Indicadores. 3. Saúde ambiental. I. Battisti, Iara Denise Endruweit, orient. II. Kaszubowski, Erikson, co-orient. III. Universidade Federal da Fronteira Sul. IV. Título.

**ALEXANDRE LUIZ SCHÄFFER**

**ESTUDO DE INDICADORES DE AMBIENTE E SAÚDE UTILIZANDO A  
METODOLOGIA DE MODELOS MULTINÍVEIS NAS MICRORREGIÕES DO RIO  
GRANDE DO SUL**

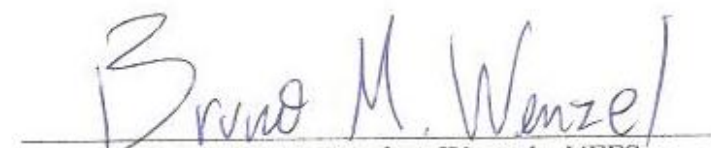
Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Ambiental e Sanitária da Universidade Federal da Fronteira Sul como requisito para obtenção do título de Engenheiro Ambiental e Sanitário.

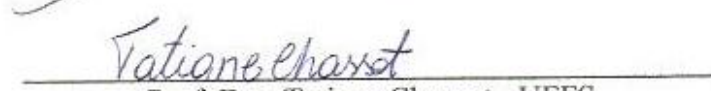
Orientadora: Dra. Iara Denise Endruweit Battisti

Coorientador: Dr. Erikson Kaszubowski

Este trabalho de conclusão de curso foi defendido e aprovado pela banca em 09 / 12 / 2016.

BANCA EXAMINADORA

  
Prof. Dr. Bruno München Wenzel - UFFS

  
Prof. Dra. Tatiane Chassot - UFFS

Dedico este trabalho a todos que de alguma forma contribuíram em minha evolução acadêmica.

## **AGRADECIMENTOS**

Primeiramente agradeço a Deus pela vida, saúde e ânimo para encarar as dificuldades encontradas durante o desenvolvimento de toda a graduação.

Em segundo lugar e não menos importante, aos meus pais Gelásio e Lourdes, como também a minha irmã Silvana, pelo apoio incondicional a mim fornecido. Sem estas pessoas, não teria encontrado forças para prosseguir meus estudos.

Agradeço também a minha orientadora Iara Denise Endruweit Battisti e ao coorientador Erikson Kaszubowski pelas horas dedicadas a mim, na troca de conhecimentos e informações, que por fim compuseram este trabalho.

Aos colegas do curso e pessoas as quais desenvolvi amizades durante a graduação, encarando juntos as mais diversas dificuldades enfrentadas pelo curso. Haverão eternas lembranças das amizades que se formaram no início das aulas no CETRAN e antigo Seminário São José.

A UFFS e a FAPERGS pelo apoio concedido durante grande parte da graduação em atividades de ensino e pesquisa. Muito obrigado! Sem este apoio, jamais conseguira o título de bacharel em Engenharia Ambiental.

Enfim, a todos os familiares e pessoas que de alguma forma contribuíram no processo de minha formação acadêmica.

## RESUMO

As modificações ambientais influenciam diretamente na qualidade de vida das pessoas. Em razão disto, é pertinente estruturar um sistema de indicadores que permita monitorar as condições ambientais que podem afetar a saúde pública. Assim, o objetivo deste estudo foi construir e analisar indicadores de saúde ambiental para as microrregiões do Estado do Rio Grande do Sul (RS), denominadas Conselhos Regionais de Desenvolvimento (COREDEs), verificando a relação de desfechos de morbidade e mortalidade com preditores demográficos, socioeconômicos e de cobertura por serviços de saúde e saneamento, os quais poderão servir como monitoramento de agravos a saúde relacionada ao ambiente. Os procedimentos metodológicos contemplaram: revisão bibliográfica da relação entre indicadores de morbidades e mortalidades (desfechos) com demais indicadores de saúde ambiental (preditores); pesquisas em sistemas de informação que disponibilizam dados de saúde, ambiente e saúde ambiental; estudo da modelagem multinível; coleta de dados nos sistemas de informação; e, ajuste e interpretação dos modelos multiníveis. Utilizaram-se modelos de regressão multinível adotando-se os municípios como unidade de análise de primeiro nível e as microrregiões do RS como de segundo nível. Foram ajustados quatro modelos com os seguintes desfechos: coeficientes de morbidade (casos/1.000 habitantes) e mortalidade (casos/100.000 habitantes) para doenças infecciosas e parasitárias (modelo 5 e modelo 6, respectivamente) e, coeficientes de morbidade (casos/10.000 habitantes) e mortalidade (casos/100.000 habitantes) para doenças diarreicas e gastroenterite (modelo 7 e modelo 8, respectivamente). Dentre os principais resultados, tem-se que o modelo 5 final, foi composto por três preditores, sendo um preditor com sinal inverso do esperado ao desfecho. O modelo 6 final, com 16 preditores, sendo oito preditores com sinal esperado ao desfecho. O modelo 7 final compôs-se de cinco preditores com sinal inverso ao desfecho esperado. Para o modelo 8 final, tem-se dez preditores, com seis preditores com sinal inverso do esperado ao desfecho. Por fim, como consideração final, as análises mostraram a aplicabilidade da técnica de modelagem multinível para investigar e monitorar relações entre indicadores de saúde e ambiente. Sugere-se explorar outros preditores que possam estar influenciando os quatro modelos ajustados, no sentido de explicar os sinais inversos encontrados.

Palavras-chave: Regressão multinível. Indicadores. Saúde ambiental.

## ABSTRACT

Environmental modifications directly influence people's quality of life. It is appropriate to structure a system of indicators to monitor environmental conditions that may affect public health. Thus, the objective of this study was to construct and analyze environmental health indicators for the micro-regions of the State of Rio Grande do Sul (RS), called Conselho Regional de Desenvolvimento (COREDEs), verifying the relationship of morbidity and mortality with demographic, socioeconomics and coverage by health and sanitation services, which may serve as a monitoring of health-related environmental damage. The methodological procedures included: bibliographic review of the relationship between morbidity and mortality indicators (outcomes) with other environmental health indicators (predictors); research in information systems that provide data on health, environment and environmental health; multilevel modeling study; search of data in information systems; and, adjustment and interpretation of multilevel models. Multilevel regression models were used adopting the municipalities as first level analysis unit and the RS micro-regions as second level. Four models with the following outcomes were adjusted: morbidity coefficients (cases / 1,000 inhabitants) and mortality (cases / 100,000 inhabitants) for infectious and parasitic diseases (model 5 and model 6, respectively) and morbidity coefficients (cases / 10,000 inhabitants) and mortality (cases / 100,000 inhabitants) for diarrheal diseases and gastroenteritis (model 7 and model 8, respectively). The results show that the final model 5 was composed of three predictors, being a predictor with an inverse signal of the one expected to the outcome. The final model 6, with 16 predictors, being eight predictors with signal expected to the outcome. The final model 7 consisted of five predictors with the opposite signal to the expected outcome. The final model 8 has ten predictors, with six predictors with the opposite sign of the expected outcome. As conclusion, the analyzes showed the applicability of the multilevel modeling technique to investigate and monitor relationships between health and environment indicators. It is suggested to explore other predictors that may be influencing the four adjusted models, in the sense of explaining the inverse signals found.

**Keywords:** Multilevel Regression. Indicator. Environmental Health.



## LISTA DE FIGURAS

Figura 1 - Inter-relação entre os três grupos de mudanças ambientais e seus potenciais impactos que podem afetar grupos populacionais vulneráveis .....	20
Figura 2 - O espaço dos indicadores de saúde ambiental .....	23
Figura 3 - Etapas para construção de indicadores .....	25
Figura 4 - Distribuição dos COREDEs no Estado do RS.....	34
Figura 5 - Distribuição das RFs no Estado do RS .....	35
Figura 6 - Estrutura de dados para o modelo hierárquico do estudo de Heimerdinger (2011) .....	40
Figura 7 - Interface gráfica do software R (3.2.5) .....	43
Figura 8 - Interface gráfica do programa RStudio.....	44
Figura 9 – Desfechos dos modelos de regressão linear múltipla e modelos multiníveis ....	52
Figura 10 - Gráficos para análise de resíduos em regressão.....	56
Figura 11 - Preditores com menos de 20% de dados faltantes para a construção dos modelos de regressão linear múltipla .....	63
Figura 12 - Regressão linear simples entre “Coeficiente de morbidade por doenças infecciosas e parasitárias” e “Porcentagem da população com abastecimento de água e esgotamento sanitários inadequados”, COREDEs do RS (2000 e 2010).....	65
Figura 13 - Regressão linear simples entre “Coeficiente de mortalidade por doenças infecciosas e parasitárias” e “Porcentagem de pessoas acima de 18 anos desocupadas”, COREDEs do RS (2000 e 2010) .....	67
Figura 14 - Regressão linear simples entre “Coeficiente de mortalidade por doenças diarreicas e gastroenterite” e “Porcentagem de pessoas vulneráveis a pobreza”, COREDEs do RS (2000 e 2010) .....	68
Figura 15 - Regressão linear simples ente “Coeficiente de morbidade por doenças diarreicas e gastroenterite” e a “Expectativa de vida ao nascer”, COREDEs do RS (2000 e 2010) .....	69
Figura 16 - Preditores com menos de 20% de dados faltantes para construção dos modelos multiníveis.....	70
Figura 17 - Gráfico dos resíduos para modelo 5 com escala logarítmica (a) e original (b)	71

Figura 18 - Gráfico dos resíduos para o modelo 6 com função logarítmica e base de dados integral (a); função original e base de dados integral (b); função original e base de dados sem zeros (c); função logarítmica e base de dados sem zeros (d)..... 73

Figura 19 - Gráfico dos resíduos para modelo 7 com função original e base de dados integral (a); com função logarítmica e com base de dados integral (b); com função original e base de dados sem zeros (c); com função logarítmica e base de dados sem zeros (d)..... 78

Figura 20 - Gráfico dos resíduos para o modelo 8 com função original e base de dados integral (a); com função logarítmica e base de dados integral (b); com função original e base de dados sem zeros (c); com função logarítmica e base de dados sem zeros (d)..... 81

## LISTA DE QUADROS

Quadro 1 - Sistemas de informação no Brasil .....	26
Quadro 2 - Fonte dos dados disponíveis no sistema DATASUS .....	28
Quadro 3 - Classificação dos dados disponíveis no sistema de informação IPEADATA ..	29
Quadro 4 - Fonte dos dados disponíveis no sistema FEEDADOS.....	31
Quadro 5 - Número de municípios por COREDE, Região Funcional e respectivo número de habitantes em 2010 no Estado do RS .....	35
Quadro 6 - Sistemas de informação estudados .....	45
Quadro 7 - Tipo de dado obtido no DATASUS e respectiva descrição .....	45
Quadro 8 - Metodologia utilizada no cálculo dos indicadores de saúde ambiental para o DATASUS.....	47
Quadro 9 - Tipo de dado obtido no ATLASBRASIL e respectiva descrição .....	48
Quadro 10 - Tipo de dado do Atlas da Vulnerabilidade Social e respectiva descrição.....	49
Quadro 11 - Tipo de dado do IPEADATA e respectiva descrição.....	50
Quadro 12 - Tipo de dado da FEEDADOS e respectiva descrição .....	51
Quadro 13 - Variáveis não utilizadas como preditoras nos modelos multiníveis .....	57

## LISTA DE TABELAS

Tabela 1 - Regressão linear multivariada para o modelo 1 final dos COREDEs do Estado do RS nos anos 2000 e 2010.....	64
Tabela 2 - Regressão linear multivariada para o modelo 2 final dos COREDEs do Estado do RS nos anos 2000 e 2010.....	66
Tabela 3 - Regressão linear multivariada para o modelo 3 final dos COREDEs do Estado do RS nos anos 2000 e 2010.....	67
Tabela 4 - Regressão linear multivariada para o modelo 4 final dos COREDEs do Estado do RS nos anos 2000 e 2010.....	69
Tabela 5 - Valores de AIC para diferentes equações ajustadas no modelo 5 .....	71
Tabela 6 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 5 final dos COREDEs do Estado do RS nos anos 2000 e 2010 .....	72
Tabela 7 - Valores de AIC para diferentes equações ajustadas no modelo 6.....	74
Tabela 8 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 6 final para os COREDEs do Estado do RS, anos 2000 e 2010 .....	76
Tabela 9 - Valores de AIC para diferentes equações ajustadas no modelo 7.....	78
Tabela 10 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 7 final para os COREDEs do Estado do RS, anos 2000 e 2010 .....	80
Tabela 11 - Valores de AIC para diferentes equações ajustadas no modelo 8.....	82
Tabela 12 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 8 final para os COREDEs do Estado do RS, anos 2000 e 2010 .....	82

## LISTA DE SIGLAS

COREDE	Conselho Regional de Desenvolvimento
DATASUS	Departamento de Informática do SUS
IBGE	Instituto Brasileiro de Geografia e Estatística
IPEA	Instituto de Pesquisa Econômica e Aplicada
OMS	Organização Mundial da Saúde
PIB	Produto Interno Bruto
PIB <i>per capita</i>	Produto Interno Bruto Por cada Pessoa
RF	Região Funcional
RS	Rio Grande do Sul
SUS	Sistema Único de Saúde

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>15</b>
<b>2</b>	<b>OBJETIVOS .....</b>	<b>18</b>
2.1	OBJETIVO GERAL.....	18
2.2	OBJETIVOS ESPECÍFICOS .....	18
<b>3</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>19</b>
3.1	RELAÇÃO SAÚDE E AMBIENTE.....	19
3.2	SAÚDE AMBIENTAL .....	21
3.3	INDICADORES DE SAÚDE AMBIENTAL.....	22
3.4	DADOS, ÍNDICES E INDICADORES .....	24
3.5	SISTEMAS DE INFORMAÇÃO .....	25
<b>3.5.1</b>	<b>Sistemas de Informação em saúde .....</b>	<b>27</b>
3.5.1.1	DATASUS.....	27
3.5.1.2	IPEA .....	29
3.5.1.3	IBGE.....	30
3.5.1.4	FEEDADOS .....	30
3.6	COREDES E RFS NO ESTADO DO RIO GRANDE DO SUL .....	33
3.7	MODELOS ESTATÍSTICOS .....	37
<b>3.7.1</b>	<b>Modelagem tradicional .....</b>	<b>37</b>
<b>3.7.2</b>	<b>Modelagem multinível.....</b>	<b>38</b>
3.7.2.1	Variáveis de agrupamento .....	39
3.7.2.2	Preditores do primeiro nível: efeito fixo e aleatório.....	40
3.7.2.3	Preditores do segundo nível .....	41
<b>4</b>	<b>METODOLOGIA .....</b>	<b>43</b>
4.1	SOFTWARE R (3.2.5) .....	43
4.2	COLETA DE DADOS NOS SISTEMAS DE INFORMAÇÃO.....	44
<b>4.2.1</b>	<b>DATASUS .....</b>	<b>45</b>
<b>4.2.2</b>	<b>ATLAS BRASIL.....</b>	<b>48</b>
<b>4.2.3</b>	<b>ATLAS DA VULNERABILIDADE SOCIAL .....</b>	<b>49</b>
<b>4.2.4</b>	<b>IPEADATA .....</b>	<b>50</b>
<b>4.2.5</b>	<b>FEEDADOS .....</b>	<b>51</b>
4.3	DESFECHOS .....	51
4.4	ESTRATÉGIAS DE MODELAGEM.....	52
<b>4.4.1</b>	<b>Modelos de regressão linear múltipla.....</b>	<b>52</b>
4.4.1.1	ETAPA 1: Elaboração dos modelos.....	52
4.4.1.2	ETAPA 2: Definição do preditor mais significativo .....	53
<b>4.4.2</b>	<b>Estratégias de modelagem multinível.....</b>	<b>54</b>
4.4.2.1	ETAPA 1: Remoção de <i>missing</i> superior a 20% .....	54
4.4.2.2	ETAPA 2: Extração da Região e Ano .....	55
4.4.2.3	ETAPA 3: Extração da variável desfecho de interesse .....	55
4.4.2.4	ETAPA 4: Verificação do gráfico de resíduos .....	56
4.4.2.5	ETAPA 5: Retirada das variáveis que não podem ser utilizadas como preditoras .....	56
4.4.2.6	ETAPA 6 (alternativa): Exclusão de zeros na base de dados.....	57
4.4.2.7	ETAPA 7: Função para a busca do melhor modelo .....	58
4.4.2.8	ETAPA 8: Resultado final dos AICs.....	62
4.4.2.9	ETAPA 9: Busca pelo melhor modelo .....	62

4.4.2.10	ETAPA 10: Modelo final .....	62
<b>5</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>63</b>
5.1	MODELOS DE REGRESSÃO LINEAR MÚLTIPLA - MODELAGEM TRADICIONAL.....	63
5.1.1	Modelo para o desfecho “Coeficiente de morbidade por doenças infecciosas e parasitárias” por 1.000 habitantes (modelo 1).....	64
5.1.2	Modelo para o desfecho “Coeficiente de mortalidade por doenças infecciosas e parasitárias” por 100.000 habitantes (modelo 2).....	66
5.1.3	Modelo para o desfecho “Coeficiente de mortalidade por doenças diarreicas e gastroenterite” por 100.000 habitantes (modelo 3) .....	67
5.1.4	Modelo para o “Coeficiente de morbidade por doenças diarreicas e gastroenterite” por 10.000 habitantes (modelo 4) .....	69
5.2	MODELOS MULTINÍVEIS .....	70
5.2.1	Modelo para o desfecho “Coeficiente de morbidade por doenças infecciosas e parasitárias” por 1.000 habitantes (modelo 5).....	70
5.2.2	Modelo para o desfecho “Coeficiente de mortalidade por doenças infecciosas e parasitárias” por 100.000 habitantes (modelo 6).....	73
5.2.3	Modelo para o desfecho “Coeficiente de morbidade por doenças diarreicas e gastroenterite” por 10.000 habitantes (modelo 7) .....	78
5.2.4	Modelo para o desfecho “Coeficiente de mortalidade por doenças diarreicas e gastroenterite” por 100.000 habitantes (modelo 8) .....	81
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>85</b>
	<b>REFERÊNCIAS .....</b>	<b>86</b>





## 1 INTRODUÇÃO

O ser humano está exposto a diversos fatores de risco, sejam eles presentes na água, no ar, no solo ou em alimentos contaminados, podendo afetar a população de forma ocupacional ou não ocupacional, como também, de maneira aguda ou crônica. Ribeiro (2004) aponta que o grande número de fatores ambientais, os quais afetam ou podem afetar a saúde humana, implicam em uma complexidade de ações necessárias, com o objetivo de melhorar os fatores ambientais determinantes da saúde humana.

Freitas (2002 apud ANDREAZZI et al., 2007) cita que após a Revolução Industrial, a população mundial mais que triplicou, a economia teve um acréscimo de 20 vezes, o consumo de combustíveis fósseis aumentou 30 vezes e a produção industrial cresceu num nível exorbitante (cerca de 50 vezes). Como consequência, ocorreu uma grande degradação do ambiente e da saúde da população.

Instituições como a Organização Mundial da Saúde (OMS) estimam que atualmente aproximadamente 1/3 das doenças que atingem as populações, tem origem de fatores ambientes, onde aproximadamente 40% dessas doenças atingem pessoas em estágio infantil (CASTRO et al., 2003). Outros autores estimam que 24% da carga total de doenças e 23% de todos os óbitos, possam ser atribuídos ao meio ambiente, afetando consideravelmente crianças de 0 a 14 anos (FRANCO NETTO et al., 2009).

Prüss et al. (2002 apud COSTA et al., 2005) apresenta em seus estudos (nível mundial) que a diarreia é responsável por 4,3% dos anos de vida perdidos e 88% desta categoria de doença está diretamente relacionada ao abastecimento de água, esgotamento sanitário e higiene inadequados, apresentando a maior concentração em crianças de países em desenvolvimento, caracterizadas por situação extremamente vulnerável ao saneamento básico.

Contudo, nos últimos 20 anos de atuação do Sistema Único de Saúde (SUS), a saúde ambiental deve ser compreendida como uma área recente nos programas governamentais, a qual retoma o debate sobre os determinantes sociais da saúde, expressos nos relatórios publicados pela OMS e Comissão Nacional sobre Determinantes Sociais da Saúde, sendo um dos desafios enfrentados pela saúde ambiental, o de definir e construir um conjunto de indicadores que propiciem subsídios para os processos de tomada de decisão no planejamento de políticas públicas que incorporem a compreensão de um quadro socioeconômico da população analisada (FRANCO NETTO et al., 2009).

Assim, percebe-se a importância de um sistema de informação que é essencial para um bom planejamento dos gestores, bem como para a transparência dos recursos públicos e um controle social, fazendo com que a tomada de decisões, em qualquer instância (municipal, estadual e federal), tenha como base, fatores relatados a partir de dados da realidade existente no local (BRASIL, 2011).

Porém, para uma correta avaliação dos efeitos gerados pelo saneamento ambiental sobre a morbi-mortalidade infantil, depende-se consideravelmente de outros fatores, tais como, demográficos, socioeconômicos, de cobertura por serviços de saúde e saneamento, tendo a observância na realização destes estudos, a consideração de várias questões metodológicas, evitando uma interpretação errônea da realidade estudada (TEIXEIRA; PUNGIRUM, 2005).

Em razão disto, a epidemiologia se apresenta como instrumento capaz de auxiliar a tomada de decisões em todas as esferas governamentais (TEIXEIRA, 2012), de tal forma que os indicadores em saúde ambiental devem subsidiar as atividades de planejamento e formulação de políticas públicas, bem como o monitoramento das condições de vida e bem-estar das pessoas (BRASIL, 2011).

Oliveira e Faria (2008) afirmam em seus estudos que a construção de indicadores de saúde ambiental constitui-se como uma oportunidade para subsidiar processos para definição de estratégias de prevenção e controle dos riscos, bem como, para a promoção da saúde, utilizando-se para isto, a análise dos resultados de estudos epidemiológicos que demonstrem os efeitos do meio ambiente sobre a saúde das pessoas.

Embora os indicadores sejam uma representação simplificada da realidade, sua construção, na maioria das vezes, não é tarefa simples. Sua complexidade faz com que eles sejam desenvolvidos com base em termos bem definidos de acordo com a escolha precisa dos dados necessários para sua correta construção e interpretação, necessitando contemplar algumas propriedades desejáveis, tais como, relevância social, validade, confiabilidade, cobertura e sensibilidade na análise das informações prestadas (BRASIL, 2011).

Uma estratégia para avaliar as relações entre variáveis ambientais e de saúde, permitindo a identificação de indicadores relevantes, é a análise baseada em modelos multiníveis. Esta técnica possibilita examinar grupos e indivíduos dentro desses grupos (simultaneamente), podendo ser a variável resposta medida a nível individual e as variáveis explicativas medidas em níveis de indivíduos ou grupos aos quais pertencem (SNIJDERS;

BOSKER, 1999 apud HEIMERDINGER, 2011). O desenvolvimento destes modelos tem viabilizado a análise de estudos que integrem indivíduos dentro de grupos com mesmo contexto social, examinando os efeitos combinados, tanto das variáveis individuais, como das variáveis relacionadas em grupo (MENDONÇA, 2001 apud MORAES, 2007).

Aplicam-se diferentes metodologias para se selecionar modelos aos desfechos estudados, tais como  $C_p$  de Mallows, Regressão Stepwise, Critério de Informação Akaike (AIC), Critério de Informação Bayesiano (BIC), dentre outros (EMILIANO, 2009).

Os critérios de informação são largamente usados, como forma de auxiliar na escolha da ordem adequada do modelo. Dentre os critérios de informação, o mais conhecido, e ainda largamente utilizado, é o Critério de Informação de Akaike (AIC – Akaike Information Criterion), em que se baseia na divergência Kullback-Leibler, que é uma medida de quanta informação é “perdida” ao tentar representar um conjunto T de medidas utilizando uma base conhecida L. Porém, como não se conhece o modelo que gera T, não se pode comparar o modelo obtido ao modelo “ideal” T. Entretanto, Akaike formulou um meio de fazer um cálculo aproximado da divergência de Kullback-Leibler com relativa simplicidade, talvez essa a razão da sua grande aceitação (SOBRAL; BARRETO, 2011).

Nos critérios AIC e BIC para cada modelo gera-se um valor, sendo que o modelo que representar o menor AIC (ou BIC) é considerado o “melhor” modelo (EMILIANO, 2009).

Com base nisto, buscou-se analisar indicadores do Estado do Rio Grande do Sul (RS), que está dividido territorialmente em 496 municípios, organizados em 28 Conselhos Regionais de Desenvolvimento (COREDEs) (ALLEBRANDT et al., 2011), tendo como objetivo, verificar a associação entre desfechos de morbidades e mortalidades através de preditores demográficos, socioeconômicos e de cobertura por serviços de saúde e saneamento, utilizando a metodologia de modelos de regressão linear múltipla e modelagem multinível, que poderão servir de monitoramento de agravos a saúde relacionada às alterações do meio ambiente.

## **2 OBJETIVOS**

### **2.1 OBJETIVO GERAL**

Verificar a associação entre indicadores de ambiente e saúde nas microrregiões do Rio Grande do Sul utilizando a metodologia de modelos multiníveis.

### **2.2 OBJETIVOS ESPECÍFICOS**

- Revisar a literatura sobre a relação entre indicadores de morbidade/mortalidade e indicadores ambientais utilizando modelagem tradicional e modelagem multinível;
- Estudar o referencial teórico sobre modelagem tradicional e modelagem multinível;
- Aplicar modelagem tradicional e multinível para verificar a relação entre os indicadores de morbidade/mortalidade e ambiente para as microrregiões do Rio Grande do Sul.

### 3 REFERENCIAL TEÓRICO

#### 3.1 RELAÇÃO SAÚDE E AMBIENTE

As concepções sobre a relação entre saneamento, vinculada à saúde das pessoas, tem se desenvolvido junto com a humanidade através das civilizações, recebendo conotações distintas de acordo com os conceitos de saúde, bem-estar e com a lógica organizacional da sociedade no decorrer do seu processo histórico de evolução e reprodução (ANDREAZZI et al., 2007).

Freitas (2002 apud ANDREAZZI et al., 2007) cita que após a Revolução Industrial, a população mundial mais que triplicou, a economia teve um acréscimo de 20 vezes, o consumo de combustíveis fósseis aumentou 30 vezes e a produção industrial cresceu num nível exorbitante (cerca de 50 vezes). Como consequência, ocorreu uma grande degradação do ambiente e da saúde da população.

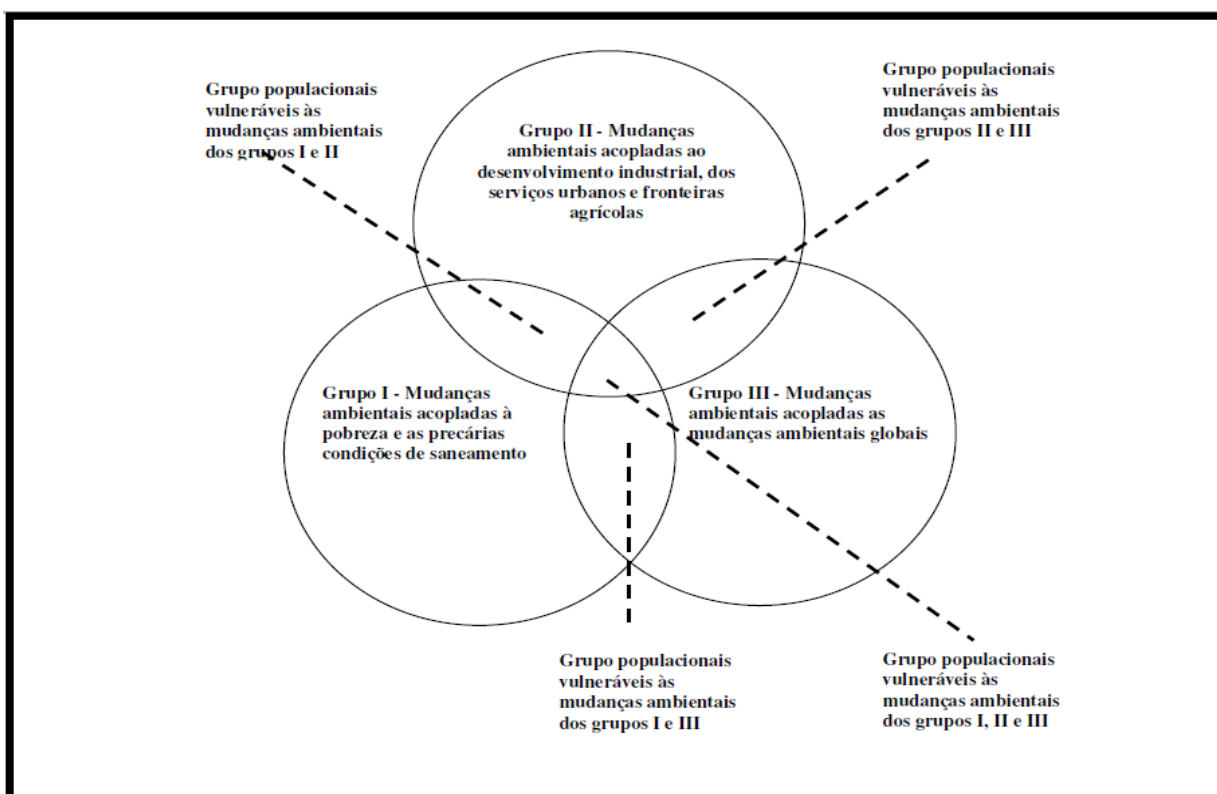
Teixeira e Guilhermino (2006), após realizarem um estudo sobre a associação entre saneamento e saúde nos Estados brasileiros, identificaram que a mortalidade infantil, a mortalidade proporcional por diarreia aguda em menores de cinco anos de idade e, a mortalidade proporcional por doenças infecciosas e parasitárias para todas as idades estavam intrinsicamente relacionadas com as deficiências na cobertura por serviços de saneamento básico ineficazes no local.

De acordo com a OMS, aproximadamente 1/3 das doenças que atingem as populações, tem origem de fatores ambientais, onde aproximadamente 40% dessas doenças atingem pessoas em estágio infantil (CASTRO et al., 2003). Outros autores estimam que 24% da carga total de doenças e 23% de todos os óbitos obtidos, possam ser atribuídos ao meio ambiente, afetando consideravelmente crianças de 0 a 14 anos (FRANCO NETTO et al., 2009).

Prüss et al. (2002 apud COSTA et al., 2005) apresenta em seus estudos (nível mundial) que a diarreia é responsável por 4,3% dos anos de vida perdidos e 88% desta categoria de doença está diretamente relacionada ao abastecimento de água, esgotamento sanitário e higiene inadequados, apresentando a maior concentração desta carga de doenças em crianças nos países em desenvolvimento, caracterizadas por situação extremamente vulnerável ao saneamento básico.

De acordo com Franco Netto et al. (2009), existem três grupos determinantes ambientais que não agem isoladamente, se encontrando combinados com determinantes sociais e mudanças ambientais, através de efeitos diretos, mediados e modulados, ao ponto que resultam em diferentes combinações e sobreposições de exposições, riscos e efeitos sobre a saúde das pessoas. A Figura 1 exemplifica os determinantes sociais e sua inter-relação.

Figura 1 - Inter-relação entre os três grupos de mudanças ambientais e seus potenciais impactos que podem afetar grupos populacionais vulneráveis



Fonte: FRANCO NETTO et al., 2009.

De acordo com Smith e Ezzati (2005), Franco Netto (2008 apud FRANCO NETTO et al., 2009), os três grupos citados na Figura 1, podem ser assim designados:

- **GRUPO I:** vinculado ao subdesenvolvimento, ou seja, ao déficit do saneamento básico ambiental, estando relacionado a estados mórbidos prevalentes e a uma forte perda da qualidade de vida das pessoas, principalmente infantil e diretamente relacionada a diarreia;
- **GRUPO II:** Contato humano direto ou indireto com o desenvolvimento industrial, dos serviços urbanos e das fronteiras agrícolas, onde seus subprodutos apresentam

toxicidade e são altamente poluentes, resultando em múltiplas consequências a saúde das comunidades e dos trabalhadores envolvidos;

- GRUPO III: Considerado emergente, resultado do macro fenômeno da globalização e da crise ambiental global, expressos na intensa urbanização, degradação dos ecossistemas e das mudanças do clima.

Teixeira e Guilhermino (2006) concluem que atualmente sabe-se que os serviços de saneamento básico são de vital importância para proteção da saúde das pessoas, bem como minimizam as consequências da pobreza, protegendo o meio ambiente em que vivemos.

### 3.2 SAÚDE AMBIENTAL

A Saúde Ambiental tem uma estreita relação com a saúde pública das pessoas, incorporando em sua concepção, graves questões como o aquecimento global, redução da camada de ozônio, desastres naturais como enchentes e tufões e ainda a proliferação de vetores como insetos e roedores, bem como, condições de moradia e saneamento básico do ser humano (RADICCHI; LEMOS, 2009).

Os mesmos autores afirmam que os problemas ambientais decorrem tanto da modernidade expansiva, quanto da miséria pela qual grande parte da população é submetida, não podendo desta forma, homogeneizar os problemas gerados, os quais podem ser classificados como de curto, médio ou longo prazo, afetando as atuais e futuras gerações.

Um dos grandes desafios enfrentados pela saúde ambiental é o de definir um conjunto de indicadores que apresentem subsídios para os processos de tomada de decisões, bem como um planejamento de políticas públicas não só centralizadas nos efeitos sobre a saúde, mas que também que incorporem a compreensão de um quadro socioeconômico e possibilitem avançar em ações intersetoriais (FRANCO NETTO et al., 2009).

De acordo com a OMS, saúde ambiental se faz como a parte da saúde pública que se ocupa das formas de vida, como também, das substâncias e das condições do entorno do homem, as quais podem exercer alguma influencia sobre a saúde e o bem-estar coletivo das pessoas (BRASIL, 2006).

Outras instituições, como o Ministério da Saúde, definem saúde ambiental como a área da saúde pública que potencialmente afeta o conhecimento científico e a formulação de políticas públicas, as quais estão diretamente relacionadas à saúde humana e aos fatores do

meio ambiente natural e antrópico, influenciando na qualidade de vida do ser humano (BRASIL, 2005). Brasil (2006) também observa que os serviços de saúde ambiental são promotores do melhoramento dos parâmetros ambientais, motivados pela aplicação de tecnologias e comportamentos favoráveis a saúde das pessoas e ao meio ambiente físico.

Ao analisar a relação dos atuais níveis de desenvolvimento e seus impactos sobre a natureza e a saúde das pessoas, é urgente a estruturação de um sistema de indicadores que possam avaliar as condições ambientais do meio ambiente, gerando um maior planejamento, execução e avaliação de ações governamentais em medidas preventivas ao combate a doenças e outros males causados a população (BORJA; MORAES, 2003 apud CALIJURI et al., 2009).

Oliveira e Faria (2008) afirmam em seus estudos que a construção de indicadores de saúde ambiental constituem-se como uma oportunidade para subsidiar processos para definição de estratégias de prevenção e controle dos riscos, bem como, para a promoção da saúde, utilizando-se para isto, a análise dos resultados de estudos epidemiológicos que demonstrem os efeitos do meio ambiente sobre a saúde das pessoas.

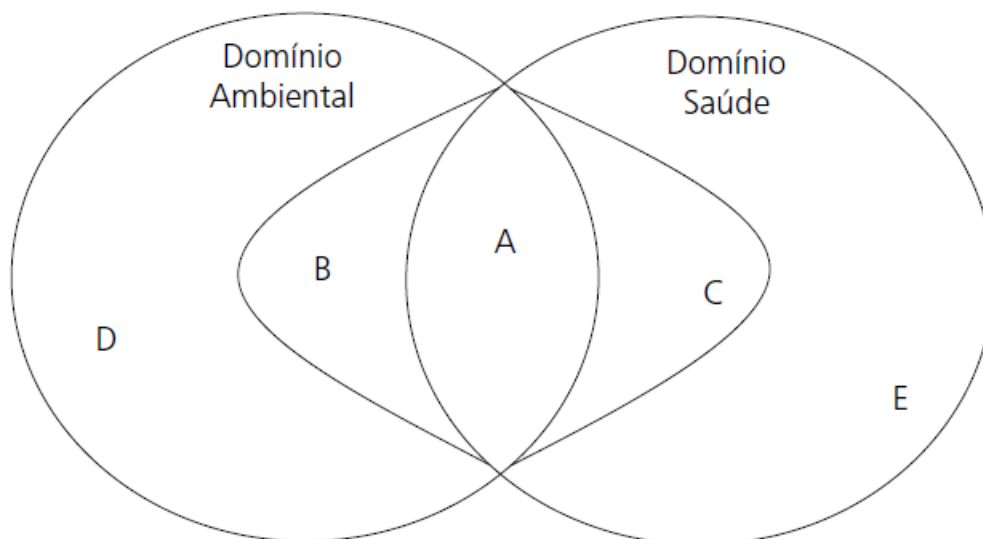
Contudo, nos últimos 20 anos de atuação do SUS, a saúde ambiental deve ser compreendida como uma área recente nos programas governamentais, a qual retoma o debate sobre os determinantes sociais da saúde, expressos nos relatórios publicados tanto pela OMS, como pela Comissão Nacional sobre Determinantes Sociais da Saúde, sendo um dos desafios enfrentados pela saúde ambiental, o de definir e construir um conjunto de indicadores que propiciem subsídios para os processos de tomada de decisão no planejamento de políticas públicas que incorporem a compreensão de um quadro socioeconômico da população analisada (FRANCO NETTO et al., 2009).

### 3.3 INDICADORES DE SAÚDE AMBIENTAL

Os indicadores de saúde ambiental podem ser definidos através de uma relação estreita entre o ambiente e a saúde das pessoas, ou também, entre um indicador ambiental e um indicador de saúde, como apresentado na Figura 2.



Figura 2 - O espaço dos indicadores de saúde ambiental



Fonte: BRASIL, 2011, adaptado de CORVALÁN; BRIGGS; KJELLSTROM, 2000.

Nota: A=Indicadores de saúde ambiental; B= indicadores ambientais de potenciais impactos à saúde humana; C= indicadores de saúde com prováveis causas ambientais; D= indicadores ambientais bem definidos; E= indicadores de saúde bem definidos.

O grande número de fatores ambientais que afetam ou podem afetar a saúde das pessoas, implicam em uma complexidade de ações necessárias para melhorar os fatores ambientais, os quais estão diretamente relacionados à saúde das populações (RIBEIRO, 2004).

Sausen (2014), utilizando o modelo Força-Motriz, Pressão, Situação, Exposição, Efeito, Ação (FPSEEA), verificou que os indicadores de saúde ambiental para a região das Missões, RS, apresentaram taxas com intensidade negativa maior em 2000, quando comparadas ao ano de 2010, afetando positivamente a saúde das pessoas residentes no local. Com o mesmo estudo, se observou menores taxas nos indicadores de morbidade e mortalidade, porém, quando os resultados foram comparados a níveis de Brasil, Região Sul e RS, observaram-se indicadores com maior frequência de intensidade negativa quando comparado a Região Sul, e muito próximas das taxas apresentadas para o Brasil.

Contudo, para uma correta avaliação dos efeitos gerados pelo saneamento ambiental sobre a morbi-mortalidade infantil, depende-se consideravelmente de outros fatores, tais como, demográficos, socioeconômicos, de cobertura por serviços de saúde, etc., tendo a

observância na realização destes estudos, a consideração de várias questões metodológicas, evitando uma interpretação errônea da realidade estudada (TEIXERIA; PUNGIRUM, 2005).

Os indicadores representam muito mais do que um simples conjunto de dados. Sintetizam em termos facilmente compreensíveis e relevantes, aspectos da relação existente entre saúde e o ambiente, orientando gestores na tomada de decisões (BRASIL, 2011).

Embora os indicadores sejam uma representação simplificada da realidade, sua construção, na maioria das vezes, não é tarefa simples. Sua complexidade faz com que eles sejam desenvolvidos com base em termos bem definidos de acordo com a escolha precisa dos dados necessários para sua correta construção e interpretação, necessitando contemplar algumas propriedades desejáveis, tais como, relevância social, validade, confiabilidade, cobertura e sensibilidade na análise das informações prestadas (BRASIL, 2011).

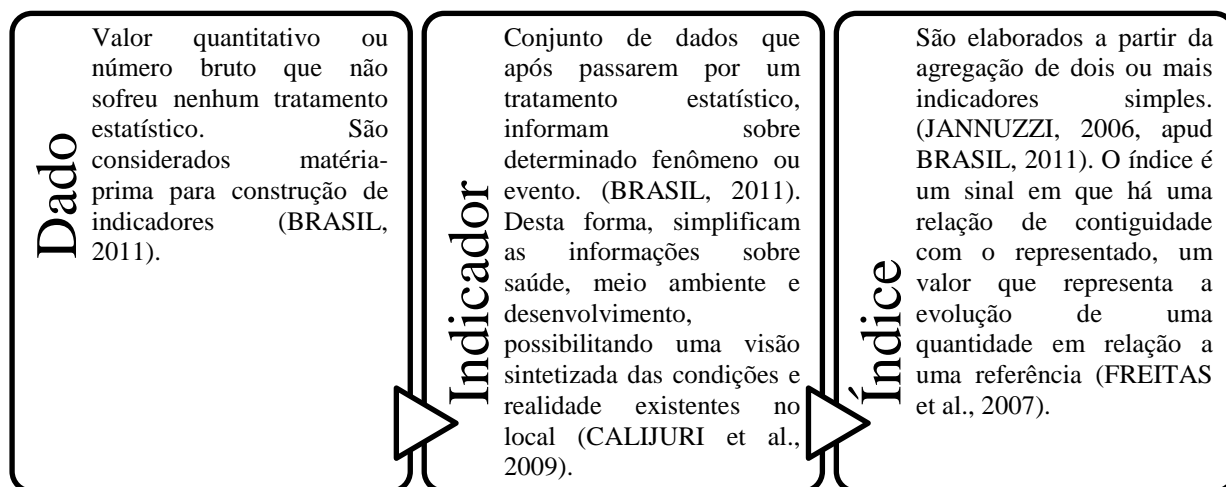
Costa et al. (2005) concluem que os indicadores de saúde ambiental, além de terem um potencial para representarem os efeitos da insuficiência das ações de saneamento sobre a saúde humana, podem se constituir como ferramenta para a vigilância e orientação de programas, na alocação de recurso públicos como os de saneamento básico.

### 3.4 DADOS, ÍNDICES E INDICADORES

Os indicadores são uma espécie de representação da realidade local, sendo por este motivo, relevantes nos processos de gestão e tomadas de decisões por parte de governantes, que muitas vezes não dispõem de tempo suficiente para uma análise adequada e criteriosa de uma ampla gama de dados, bem como um grande número de informações a cerca de um determinado problema (SILVA, 2011).

Para a construção dos indicadores de saúde ambiental, inicialmente há a necessidade de se conhecer sua estrutura de formação. A Figura 3 indica a definição de dado, indicador e índice.

Figura 3 - Etapas para construção de indicadores



Fonte: Elaborado pelo autor, 2016. Dados extraídos de BRASIL, 2011; CALIJURI, 2009 e FREITAS, 2007.

### 3.5 SISTEMAS DE INFORMAÇÃO

Para Medronho et al. (2007), um sistema é um conjunto integrado de partes que se articulam para uma finalidade específica em comum. Sua característica fundamental é a capacidade de coligir um amplo número de registros, bem como, armazená-los e recuperá-los, correspondendo às funções dos bancos de dados.

Um sistema de informação é essencial para um bom planejamento dos gestores, bem como para a transparência dos recursos públicos e um controle social, fazendo com que a tomada de decisões, em qualquer instância (municipal, estadual e federal), tenha como base, fatores relatados a partir de dados da realidade existente no local (BRASIL, 2011).

Ainda, Brasil (2011) relata que o Censo se apresenta como a principal fonte de dados para a construção e elaboração de indicadores necessários a formulação de políticas públicas no Brasil, nos mais diversos setores, como, saúde, economia, ambiente, educação, etc., ou órgãos como o Ministério da Saúde, através do Departamento de Informática do SUS (DATASUS), produzindo dados que também alimentam o sistema brasileiro de estatísticas.

No Quadro 1 são apresentados os principais sistemas de informação públicos no Brasil e respectivos endereços eletrônicos.

Quadro 1 - Sistemas de informação no Brasil

(continua)

<b>SISTEMA DE INFORMAÇÃO</b>	<b>SIGLA</b>	<b>ENDEREÇO</b>
Departamento de Informática do SUS	DATASUS	<a href="http://www.datasus.gov.br">www.datasus.gov.br</a>
Atlas do Desenvolvimento Humano no Brasil 2013	ATLAS BRASIL	<a href="http://www.atlasbrasil.org.br">www.atlasbrasil.org.br</a>
Sistemas de Informações sobre Mortalidade	SIM	<a href="http://svs.aids.gov.br/cgiae/sim/">svs.aids.gov.br/cgiae/sim/</a>
Sistema de Informações sobre Nascidos Vivos	SINASC	<a href="http://svs.aids.gov.br/cgiae/sinasc/">svs.aids.gov.br/cgiae/sinasc/</a>
Sistema de Informações Hospitalares	SIH	<a href="http://www.datasus.gov.br/SIHD/institucional">www.datasus.gov.br/SIHD/institucional</a>
Sistema de Informações Ambulatoriais	SAI	<a href="http://sia.datasus.gov.br/principal/index.php">sia.datasus.gov.br/principal/index.php</a>
Sistema de Informações da Atenção Básica	SIAB	<a href="http://dab.saude.gov.br/portaldab/siab.php">dab.saude.gov.br/portaldab/siab.php</a>
Sistema de Informações sobre Agravos de Notificação	SINAN	<a href="http://dtr2004.saude.gov.br/sinanweb/">dtr2004.saude.gov.br/sinanweb/</a>
Sistema Nacional de Informação Tóxico-Farmacológica	SINTOX	<a href="http://www.fiocruz.br/sinitox/cgi/cgilua.exe/sys/start.htm?tpl=home">www.fiocruz.br/sinitox/cgi/cgilua.exe/sys/start.htm?tpl=home</a>
Sistema de Vigilância Alimentar e Nutricional	SISVAN	<a href="http://www.consea.sp.gov.br/noticia.php?id=124#.VDLpl_lidWC0">www.consea.sp.gov.br/noticia.php?id=124#.VDLpl_lidWC0</a>
Indicadores e Dados Básicos para a Saúde da Rede Interagencial de Informação para a Saúde	RIPSA	<a href="http://www.ripsa.org.br/php/index.php">www.ripsa.org.br/php/index.php</a>
Relação Anual de Informações Sociais	RAIS	<a href="http://www.rais.gov.br">www.rais.gov.br</a>
Instituto de Pesquisa Econômica e Aplicada	IPEA	<a href="http://www.ipea.gov.br/portal">www.ipea.gov.br/portal</a>
Portal Nacional de Licenciamento Ambiental	PNLA	<a href="http://integradorpnla.mma.gov.br/integrador/RelatorioBuscaLicenca.jsp">integradorpnla.mma.gov.br/integrador/RelatorioBuscaLicenca.jsp</a>
Portal Brasileiro sobre Biodiversidade	PORTALBIO	<a href="http://portalbio.wordpress.com/">portalbio.wordpress.com/</a>

(continuação)

<b>SISTEMA DE INFORMAÇÃO</b>	<b>SIGLA</b>	<b>ENDEREÇO</b>
Portal Nacional de Gestão Florestal	PNGF	<a href="http://www.florestal.gov.br/pngf">www.florestal.gov.br/pngf</a>
Sistema de Informações Ambientais	SISAM	<a href="http://sisam.cptec.inpe.br/sisam/">sisam.cptec.inpe.br/sisam/</a>
Painel de Informações em Saúde Ambiental e Saúde do Trabalhador	PISAST	<a href="http://www.saude.gov.br/svs/pisast">www.saude.gov.br/svs/pisast</a>
Instituto Brasileiro de Geografia e Estatística	IBGE	<a href="http://www.ibge.gov.br">www.ibge.gov.br</a>

Fonte: Elaborado pelo autor, 2016.

### 3.5.1 Sistemas de Informação em Saúde

De acordo com Zanini (2007), os sistemas de informação em saúde são compreendidos por um conjunto de subsistemas de informação de natureza demográfica, epidemiológica, administrativa e gerencial, necessários para o estudo e gestão dos bens e serviços da saúde das pessoas. A autora complementa que a presença de sistemas de informação bem desenvolvidos, constitui-se com um indicativo de melhor estruturação dos serviços de vigilância em saúde e consequentemente, uma maior organização de serviços de atenção e qualidade nos atendimentos prestados aos usuários.

A seguir, apresentam-se alguns dos sistemas de informação em saúde disponíveis no Brasil.

#### 3.5.1.1 DATASUS

O Departamento de Informática do SUS (DATASUS) é um sistema de informação em que são alimentadas informações de diversos sistemas, constituindo-se desta forma, um bom meio de pesquisa para coleta de dados.

De acordo com o portal do Ministério da Saúde (<http://www.datasus.gov.br>) o DATASUS disponibiliza informações que podem servir de meio para subsidiar análises objetivas da situação sanitária do local, tomadas de decisões baseadas em evidências reais, bem como a elaboração de programas de ações à saúde das pessoas.

Ainda, de acordo com o mesmo autor, o TABNET (sistema de informação do DATASUS) é considerado como um tabulador genérico de domínio público, o qual permite organizar os dados de maneira rápida conforme a consulta, sendo desenvolvido para gerar informações das bases de dados do Sistema Único de Saúde.

O TABNET apresenta diversas informações, tais como:

- Informações sobre mortalidade, nascidos vivos, epidemiológicas, morbidades, indicadores de saúde, assistência à saúde, demográficas e socioeconômicas, inquéritos e pesquisas e cadastramento da rede assistencial;
- Permite a escolha da abrangência geográfica desejada: municípios, unidades de federação, Estados;
- Permite a seleção de variáveis, tais como, sexo, faixa etária, tipo de doença, entre outros;
- Permite selecionar o período desejado, com várias informações anuais ou com intervalos de dez anos (dados vinculados ao Censo do IBGE), apresentando dados disponíveis a partir do ano de 1991;
- Gera a visualização dos resultados na forma de mapas, gráficos ou tabelas bidimensionais (linhas e colunas).

A fonte dos dados disponíveis no sistema de informação DATASUS é apresentada no Quadro 2.

Quadro 2 - Fonte dos dados disponíveis no sistema DATASUS  
(continua)

CATEGORIA	FONTE DOS DADOS
Indicadores de Saúde	IDB/ Cadernos de Saúde/ pactuação
Assistência a Saúde	Sistemas hospitalares/ Ambulatoriais/ Imunizações/ Atenção Básica/ Vig. Nutricional
Epidemiológicas e Morbidade	Sistemas de morbidade/ Agravos/ Nutrição
Rede Assistencial	Cadastro nacional de Estabelecimentos de Saúde
Estatísticas vitais	Sistema de Mortalidade/ Nascidos vivos
Demográficas e Socioeconômicas	População/ PIB/ saneamento

(continuação)

<b>CATEGORIA</b>	<b>FONTE DOS DADOS</b>
Inquéritos e pesquisas	PNAD, VIGITEL / VIVA
Saúde suplementar (ANS)	Agência Nacional de saúde Suplementar
Estatísticas de Acesso ao TABNET	Estatísticas de acesso

Fonte: Dados extraídos em <http://www2.datasus.gov.br/>.

### 3.5.1.2 IPEA

De acordo com o Instituto de Pesquisa Econômica e Aplicada (IPEA) (<http://www.ipeadata.gov.br>), a instituição IPEA é considerada como uma fundação econômica pública de categoria federal, a qual está diretamente vinculada ao Ministério do Planejamento, Orçamento e Gestão, sendo que suas atividades de pesquisa fornecem suporte técnico e institucional em decorrência das ações governamentais para a contribuição na formulação de políticas públicas e programas de desenvolvimento no país.

O Quadro 3 apresenta as três categorias dos dados disponibilizados pelo sistema de informação IPEADATA.

Quadro 3 - Classificação dos dados disponíveis no sistema de informação IPEADATA

<b>CATEGORIA</b>	<b>DESCRIÇÃO</b>
Macroeconômico	Dados econômicos e financeiros do Brasil, em séries anuais, mensais e diárias na mesma unidade monetária.
Regional	Dados econômicos, demográficos e geográficos para Estados, municípios (e suas áreas mínimas comparáveis), regiões administrativas e bacias hidrográficas brasileiras.
Social	Dados e indicadores sobre a distribuição de renda, pobreza, educação, saúde, previdência social e segurança pública.

Fonte: Dados extraídos em <http://www.ipeadata.gov.br/>.

### 3.5.1.3 IBGE

De acordo com informações do próprio instituto Brasileiro de Geografia e Estatística (IBGE) (<http://www.ibge.gov.br>), o sistema disponível se constitui como o principal provedor de dados e informações do Brasil, atendendo as necessidades dos mais diversos segmentos da sociedade civil, bem como dos órgãos das esferas governamentais (federal, estadual e municipal).

Ainda, de acordo com o sistema de informação do IBGE, este fornece uma visão completa e atual do país, pelo desempenho de suas principais funções, as quais citam-se:

- produção e análises de informações estatísticas;
- coordenação e consolidação das informações estatísticas;
- produção e análise das informações geográficas;
- coordenação e consolidação das informações geográficas;
- estruturação e implantação de um sistema de informações ambientais;
- documentação e disseminação de informações;
- coordenação dos sistemas estatísticos e cartográfico nacionais.

Para que o IBGE consiga cobrir todo o país, este constitui uma rede nacional de pesquisa, composta por:

- 27 Unidades Estaduais (nas 26 capitais e 1 no distrito Federal);
- 27 Setores de Documentação e Disseminação de Informações (26 nas capitais e 1 no Distrito Federal);
- 581 Agências de Coleta de Dados nos principais municípios.

### 3.5.1.4 FEEDADOS

O sistema de informação da Fundação de Economia e Estatística do RS (FEEDADOS), disponível em <http://feedados.fee.tche.br>, reúne informações de natureza socioeconômica relativa ao Estado do Rio Grande do Sul, bem como aos seus municípios pertencentes, sendo um banco de dados dinâmico, o qual permite realizar cruzamentos entre variáveis, unidades geográficas e período de abrangência.



As informações apresentadas no sistema de informações da FEEDADOS estão disponibilizadas a partir do ano de 1990. Pode-se optar pela coleta de dados em nível de municípios, microrregiões, mesorregiões, Conselhos regionais de Desenvolvimento (COREDEs), Região Metropolitana de Porto Alegre (RMPA) e ao nível do Estado do RS. As variáveis disponibilizadas abrangem a contabilidade social, agropecuária, indústria, comércio, exportações, demografia, emprego, saúde, educação, finanças públicas, estatísticas eleitorais, dentre outras.

Alguns dados disponibilizados são oriundos da própria Fundação de Economia e Estatística do RS e outras diversas informações disponibilizadas pelo sistema são oriundas de outros sistemas de informação, apresentados no Quadro 4.

Quadro 4 - Fonte dos dados disponíveis no sistema FEEDADOS

(continua)

<b>CATEGORIA</b>	<b>FONTE DOS DADOS</b>
Vendas de combustíveis	Agência Nacional do petróleo
Instituições financeiras	Banco Central do Brasil
Comunicações	CRT Brasil Telecon; CTMR Brasil Telecon
Energia Elétrica	Companhia Estadual de Energia Elétrica AES-Sul – Distribuição Gaúcha de Energia; RGE Rio Grande Energia A.
Saúde Demografia/Óbitos	DATASUS
Caracterização do Território/Distância de Porto Alegre	Departamento Autônomo de Estradas de Rodagem
Demografia/População, Expectativa de Vida ao Nascer e Taxa de Urbanização Educação/Taxa de Analfabetismo	FEE/Núcleo de Indicadores Sociais
IDESE	FEE/Centro de Informações Estatísticas
Contabilidade Social	FEE/Núcleo de Contabilidade Social
Demografia/Expectativa de Vida ao Nascer	IDH 2000 – ONU

(continuação)

<b>CATEGORIA</b>	<b>FONTE DOS DADOS</b>
Agricultura Pecuária Demografia/Casamentos, Separações, Divórcios, População e Taxa de Urbanização Silvicultura Extração Vegetal Contabilidade Social Domicílios Particulares Permanentes Educação/Taxa de Analfabetismo	Instituto Brasileiro de Geografia e Estatística
Educação	Instituto Nacional de Estudos e Pesquisas Educacionais – INEP
Finanças Públicas/Tributos Federais	Ministério da Fazenda - Secretaria da Receita Federal
Comércio/Exportações	Secretaria de Comércio Exterior
Educação	Secretaria de Educação
Indústria de Transformação Indústria Extrativa Mineral Comércio Atacadista e Varejista (Número de Estabelecimentos, Saídas para Outros Estados, Entradas de Outros Estados e Valor Adicionado Fiscal) Finanças Públicas/Tributos Estaduais	Secretaria da Fazenda
Transportes/Veículos Registrados	Secretaria da Justiça e da Segurança - Departamento Estadual de Trânsito
Demografia/Nascidos Vivos, Óbitos e Mortalidade Infantil	Secretaria da Saúde / Coordenadoria de Informação em Saúde – CIS/SES-RS
Transportes/Aeroportos, Aeroclubes e Bases Aéreas	Secretaria de Transportes - Departamento Aeroportuário
Justiça	Superintendência dos Serviços Penitenciários
Finanças Públicas/Tributos Municipais/Receitas Arrecadadas e Despesas Realizadas	Tribunal de Contas
Representação Política	Tribunal Regional Eleitoral

Fonte: Elaborado pelo autor. Dados extraídos em <http://feedados.fee.tche.br>.

### 3.6 COREDES E RFS NO ESTADO DO RIO GRANDE DO SUL

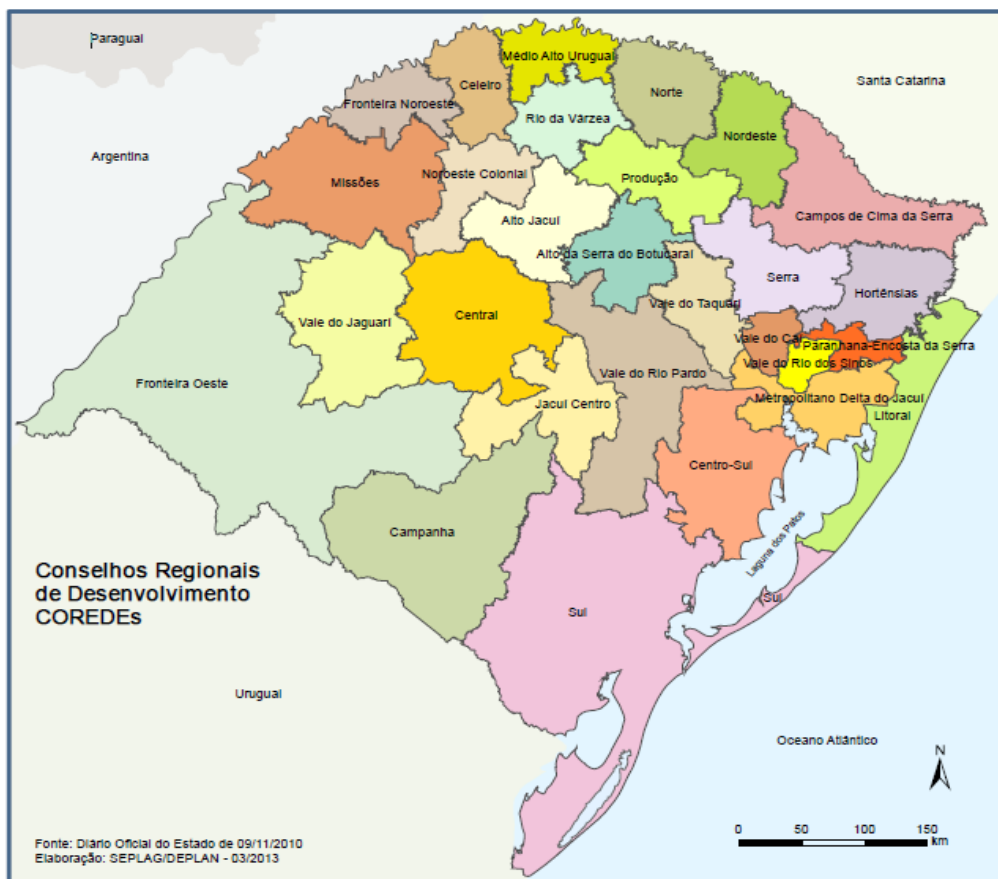
Para a construção de indicadores de saúde ambiental para microrregiões do Rio Grande do Sul, utilizou-se a classificação dos Conselhos Regionais de Desenvolvimento (COREDEs) do Estado do RS.

A estrutura organizacional em COREDEs no Estado do RS foi implantada pela lei de 17 de outubro de 1994, tendo como objetivos, promover o desenvolvimento regional, a busca de ações do governo para melhoria da qualidade de vida da população, bem como uma distribuição equitativa da renda entre as pessoas residentes no Estado do RS (ATLAS, 2013).

Inicialmente os Conselhos Regionais de Desenvolvimento eram compostos por 21 regiões. Posteriormente, em 1998, criou-se o 22º COREDE denominado Metropolitano Delta do Jacuí. No ano de 2006, criaram-se também os COREDEs Campos de Cima da Serra e Rio da Várzea e, em 2008, iniciaram-se as atividades dos COREDEs Vale do Jaguari e Celeiro (ATLAS, 2013).

Atualmente, o Rio Grande do Sul possui 28 Conselhos Regionais de Desenvolvimento (FIGURA 4).

Figura 4 - Distribuição dos COREDEs no Estado do RS



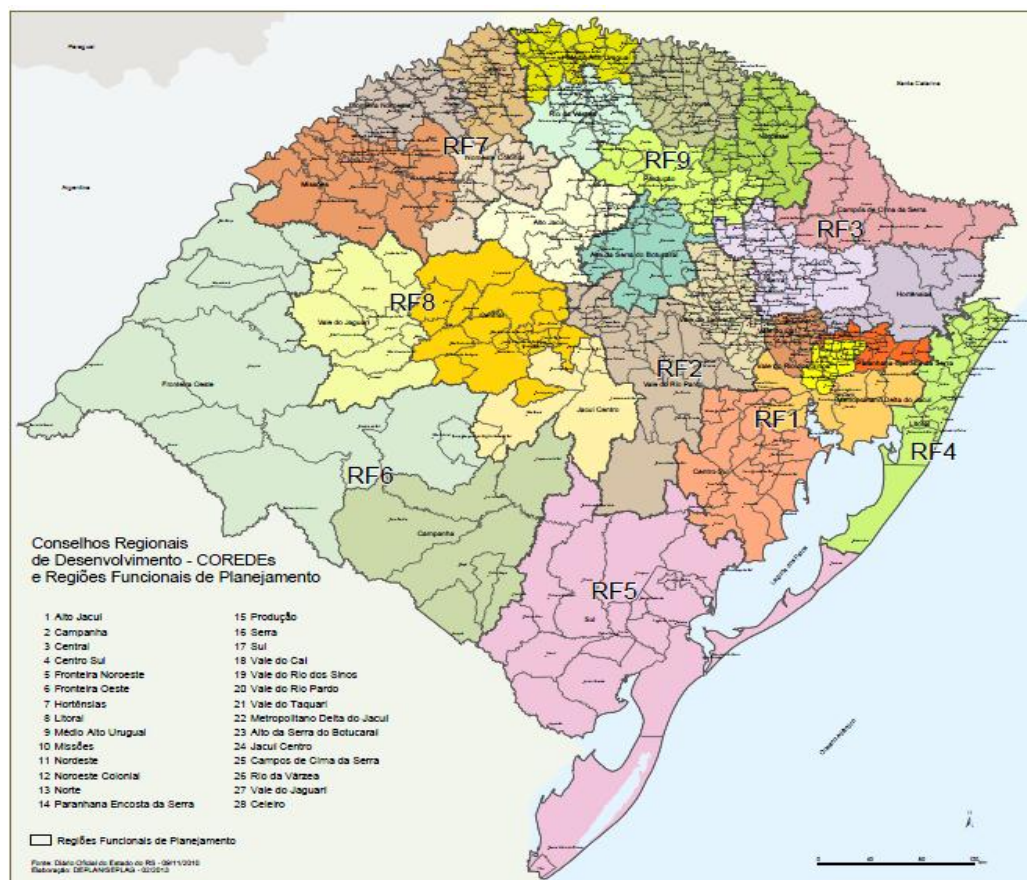
Fonte: ATLAS, 2013 (<http://www.scp.rs.gov.br/atlas>).

Os COREDEs estão agrupados em regiões denominadas Regiões Funcionais de Planejamento (RF), tendo por objetivos, a elaboração do Plano Plurianual, bem como orçamentos do Estado do RS de acordo com a característica de cada região (ATLAS, 2013).

Para o agrupamento dos municípios que compõem cada RF, foram considerados critérios com base na homogeneidade econômica, ambiental e social, para uma adequação e caracterização das populações, como emprego, saúde, educação, entre outros, para cada microrregião do Estado do RS (ATLAS, 2013).

As RFs presentes no Estado do RS são apresentadas na Figura 5.

Figura 5 - Distribuição das RFs no Estado do RS



Fonte: ATLAS, 2013 (<http://www.scp.rs.gov.br/atlas>).

No Quadro 5 apresentam-se os COREDEs, com seus respectivos números de municípios e habitantes residentes no local.

Quadro 5 - Número de municípios por COREDE, Região Funcional e respectivo número de habitantes em 2010 no Estado do RS

(continua)

REGIÃO FUNCIONAL	COREDE	Nº MUNICÍPIOS	Nº HABITANTES
RF 1	Centro Sul	17	253.461
	Paranhana Encosta da Serra	10	204.908
	Vale do Caí	19	169.580
	Vale do Rio dos Sinos	14	1.290.491
	Metropolitano Delta do Jacuí	10	2.420.262

(continuação)

<b>REGIÃO FUNCIONAL</b>	<b>COREDE</b>	<b>Nº MUNICÍPIOS</b>	<b>Nº HABITANTES</b>
RF 2	Vale do Rio Pardo	23	418.141
	Vale do Taquari	36	327.723
RF 3	Hortênsias	07	126.985
	Serra	32	862.305
	Campos de Cima da Serra	10	98.018
RF 4	Litoral	21	296.083
RF 5	Sul	22	843.206
RF 6	Campanha	07	216.269
	Fronteira oeste	13	530.150
RF 7	Fronteira Noroeste	20	198.861
	Missões	25	248.016
	Noroeste Colonial	11	166.599
	Celeiro	21	141.482
RF 8	Alto Jacuí	14	155.264
	Central	19	391.633
	Jacuí Centro	07	143.340
	Vale do Jaguarí	09	117.250
RF 9	Médio Alto Uruguai	22	148.403
	Nordeste	19	126.872
	Norte	32	221.418
	Produção	21	338.049
	Alto da Serra do Botucaraí	16	103.979
	Rio da Várzea	20	130.548

Fonte: Dados coletados em ATLAS, 2013 ([www.atlasbrasil.org.br](http://www.atlasbrasil.org.br)) e IBGE ([www.ibge.gov.br](http://www.ibge.gov.br)). Elaborado pelo autor, 2016.

Os dados do Quadro 5 referem-se ao número de municípios que compõem cada COREDE e sua respectiva RF de acordo com o Atlas de 2013. Já os dados da população são referentes às informações do Censo Demográfico de 2010 do IBGE, visto que o mesmo é realizado no período decenal, portanto, utiliza-se aqui os dados mais recentes disponíveis.

### 3.7 MODELOS ESTATÍSTICOS

#### 3.7.1 Modelagem tradicional

A modelagem de regressão é uma das técnicas estatísticas mais utilizadas para investigar a relação existente entre as variáveis estudadas (FERRÃO, 2003). No modelo de regressão,  $y_i$  representa o valor do desfecho do elemento  $i$  e  $x_i$  representa o valor do preditor. A equação que representa estas duas variáveis em estudo é dada por:

$$y_i = \alpha + \beta x_i \quad \text{Eq. 1}$$

Onde:

$\alpha$  = intercepto;

$\beta$  = coeficiente de inclinação.

A equação 1 como é um modelo de regressão linear simples ou, também chamado de modelo de regressão univariado.

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad \text{Eq. 2}$$

Onde:

$$\varepsilon_i = N(0, \sigma^2) \quad \text{Eq. 3}$$

Ou seja,  $\varepsilon_i$  são erros aleatórios de uma população normal, com média 0 e variância constante.

Os modelos de regressão tradicional se baseiam na suposição de que os indivíduos em estudo são independentes entre si com relação ao desfecho (BARROS, 2001). O mesmo autor cita como exemplo, o fato de que um indivíduo apresentar a doença de interesse não altera a probabilidade de qualquer outro indivíduo do grupo, apresentar a mesma doença.

No modelo de regressão linear múltipla considera-se duas ou mais variáveis preditoras, podendo ser expresso como:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad \text{Eq. 4}$$

Onde:

$\alpha$  = coeficiente constante;

$\beta_i$  = coeficientes associadas as variáveis preditoras  $x_{pi}$ .

### 3.7.2 Modelagem Multinível

Para a investigação epidemiológica, como também em outras áreas, é comum que indivíduos estejam organizados de maneira hierárquica, ou seja, indivíduos agrupados em unidades de nível superior, os quais podem estar agrupados em unidades dentro de outras unidades (BARROS, 2001).

O desenvolvimento de modelos multiníveis tem viabilizado a análise de estudos que integrem indivíduos dentro de grupos com mesmo contexto social, examinando os efeitos combinados, tanto das variáveis individuais, como das variáveis relacionadas em grupo (MENDONÇA, 2001, apud MORAES, 2007).

O autor descreve que esta nova abordagem não tem o objetivo de substituir, mas sim, o de complementar e enriquecer o enfoque mais dominante, buscando ir além das explicações individuais e biológicas.

A modelagem multinível é uma metodologia de análise de dados que apresentam complexos padrões de variabilidade, como por exemplo, os alunos nas salas de aulas, os funcionários em empresas, os pacientes em hospitais. Esta metodologia possibilita examinar grupos (ou amostras de grupos) e indivíduos (ou amostras de indivíduos) dentro desses grupos simultaneamente, considerando a variável desfecho medida no nível individual e as variáveis preditoras, as quais podem ser medidas no nível dos indivíduos ou nos grupos aos quais pertencem (SNIJDERS; BOSKER, 1999 apud HEIMERDINGER, 2011).

De acordo com Goldstein (2003), o modelo multinível pode ser interpretado como um sistema hierárquico de equações de regressão, possibilitando os efeitos intragrupo (efeitos individuais), como também dos efeitos entregupos (efeitos contextuais).

De acordo com Diez-Roux (1998 apud BARROS, 2001), reconhecer e levar em conta a estrutura hierárquica dos dados é importante por vários motivos, pois em muitos casos, estudar o “contexto”, ou seja, as características das unidades superiores é o principal objetivo da análise. Já em outras situações, a hierarquia dos dados é importante para não implicar na superestimação dos coeficientes presentes no modelo.

O modelo de regressão multinível incorpora naturalmente de forma parcimoniosa, a estrutura hierárquica ou de agrupamento da população em estudo, podendo tratar o intercepto e os coeficientes de inclinação como variáveis aleatórias (FERRÃO, 2003).



Um modelo de regressão multinível difere de um modelo tradicional, por ser constituído de dois componentes: um fixo e outro aleatório (SNIJDERS; BOSKERS, 1999 apud HEIMERDINGER, 2011). A parte fixa indica a magnitude das associações entre as variáveis e a parte aleatória mostra as diferenças dos grupos e as variâncias dos diferentes níveis (MERLO, 2003 apud HEIMERDINGER, 2011).

Barros (2001) cita que quando os dados são estruturados em hierarquias, as unidades do mesmo grupo raramente são independentes, isso porque as unidades em cada nível compartilham de um mesmo ambiente, como também, apresentam características semelhantes.

Ainda de acordo com Barros (2001), o fato de que as unidades compartilham o mesmo ambiente, ou serem mais semelhantes entre si, quando comparados a outras unidades, pode-se levar a uma maior similaridade com relação aos desfechos de interesse. Se isto acontecer, a suposição de independência é violada, passando então a existir correlação entre as unidades dentro de um mesmo grupo, sendo este problema ainda mais importante quando variáveis explanatórias de níveis superiores de hierarquia são de interesse, de maneira com que todas as unidades de um grupo são expostas de forma idêntica ao fator.

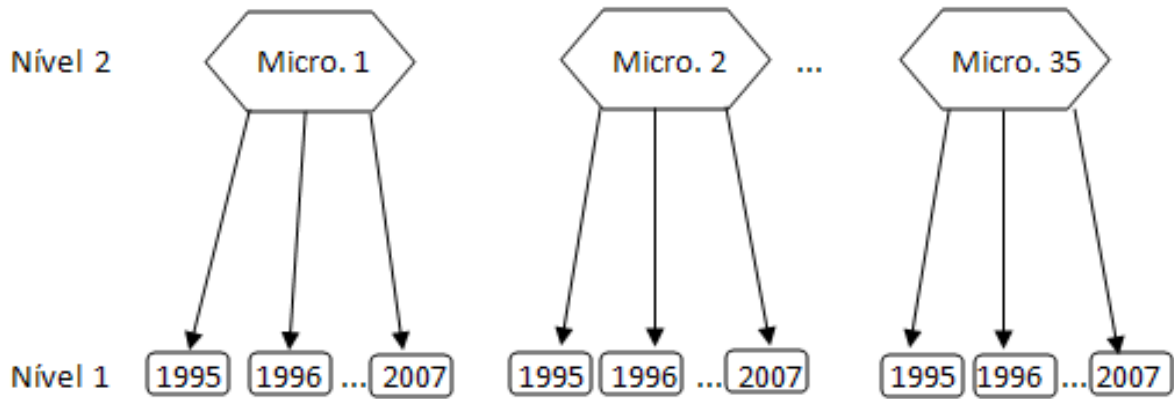
De acordo com Ferrão (2003), as vantagens da utilização do modelo de regressão multinível podem ser resumidas em quatro pontos:

- Possibilita a obtenção de estimativas eficientes dos coeficientes de regressão e de suas variâncias;
- Considera a informação sobre o agrupamento dos dados, corretos erros-padrão, intervalos de confiança, como também, testes de hipóteses;
- Permite a utilização de variáveis explicativas mensuradas em cada nível da hierarquia, permitindo uma exploração detalhada do impacto e da contribuição de cada nível para a variabilidade do desfecho;
- Permite estabelecer listas comparativas do desempenho institucional.

#### 3.7.2.1 Variáveis de agrupamento

Como a análise multinível é um estudo desenvolvido em níveis, os dados são então agrupados em grupos ou níveis. Pode-se exemplificar a estrutura do modelo multinível utilizada por Heimerdinger (2011) (Figura 6).

Figura 6 - Estrutura de dados para o modelo hierárquico do estudo de Heimerdinger (2011)



Fonte: HEIMERDINGER, 2011.

Inicialmente, o modelo mais simples a ser trabalhado em uma modelagem multinível é o Modelo Simples ou Modelo Nulo. Conhecido assim, por não apresentar variáveis preditoras ( $x_i$ ), constituindo-se apenas do intercepto e um erro aleatório associado ao modelo (Eq. 5).

$$y_{ij} = \alpha_j + \varepsilon_{ij} \quad \text{Eq. 5}$$

Onde:

$$\alpha_j = N(\alpha, \sigma^2) \quad \text{Eq. 6}$$

$$\varepsilon_{ij} = N(0, \sigma^2) \quad \text{Eq. 7}$$

Ou seja, a Equação 6 apresenta uma distribuição normal, com uma tendência central estimada pelo modelo e variância constante. Na Equação 7 observa-se que os erros aleatórios apresentam uma população normal, com média 0 e variância constante.

### 3.7.2.2 Preditores do primeiro nível: efeito fixo e aleatório

Quando utiliza-se modelagem multinível, pode-se optar por duas formas de modelar os dados: manter os preditores fixos (Efeito Fixo) ou aliar os preditores a efeitos aleatórios.

A Equação 8 apresenta um modelo ajustado com efeito fixo.

$$y_{ij} = \alpha_j + \beta_1 x_i + \dots + \beta_n x_i + \varepsilon_{ij} \quad \text{Eq. 8}$$

Onde:

$$\alpha_j = N(\alpha, \sigma^2) \quad \text{Eq. 9}$$

$$\varepsilon_{ij} = N(0, \sigma^2) \quad \text{Eq. 10}$$

Ou seja, a Equação 9 apresenta uma distribuição normal, com uma tendência central estimada pelo modelo e variância constante. Na Equação 10 observa-se que os erros aleatórios apresentam uma população normal, com média 0 e variância constante.

A equação 11 apresenta o modelo com efeitos aleatórios.

$$y_{ij} = \alpha_{j[i]} + \beta_{1j[i]}x_i + \dots + \beta_{nj[i]}x_i + \epsilon_i \quad \text{Eq. 11}$$

Onde:

$$\alpha_{j[i]} = N(\alpha, \sigma^2) \quad \text{Eq. 12}$$

$$\beta_{nj[i]} = N(\rho_n, \sigma^2) \quad \text{Eq. 13}$$

$$\epsilon_{ij} = N(0, \sigma^2) \quad \text{Eq. 14}$$

Ou seja, as Equações 12 e 13 apresentam uma distribuição normal, com uma tendência central própria estimada pelo modelo e variância constante. Na Equação 14 observa-se que os erros aleatórios apresentam uma população normal, com média 0 e variância constante.

Nos parâmetros subscritos na Equação 11, lê-se  $j$  pode ser alterado de acordo com  $i$ . No caso do presente estudo,  $i$  representa os municípios e  $j$  as microrregiões, definidas pelos COREDES.

### 3.7.2.3 Preditores do segundo nível

De acordo com Pretto (2003 apud HEIMERDINGER, 2011), para um modelo hierárquico de dois níveis, como escolas e alunos, podem-se ajustar equações de regressão separadas em cada unidade do nível, como segue:

$$y_{ij} = \alpha_j + \beta_{1j}x_i + \dots + \beta_{nj}x_i + \epsilon_i \quad \text{Eq. 14}$$

Onde:

$\alpha_j$ = intercepto da equação para cada unidade do nível 2 (neste caso, escola);

$\beta_{nj}$ = coeficiente de inclinação para cada unidade do nível 1 (neste caso, aluno);

$\epsilon_i$ = erro associado ao nível 1 (neste caso, aluno).

Onde, temos que:

$$\epsilon_i = N(0, \sigma^2) \text{ (1º nível)} \quad \text{Eq. 15}$$

$$\beta_{nj} = N(\rho_n, \sigma^2) \text{ (1º nível)} \quad \text{Eq. 16}$$

$$\alpha_j = \gamma_0 + \gamma_1 x_j + \dots + \gamma_n x_j + u_j \text{ (2º nível) – efeitos fixos} \quad \text{Eq. 17}$$

$$u_j = N(0, \sigma_2) \text{ (2º nível)} \quad \text{Eq. 18}$$

Onde:

$\gamma_0$ = intercepto para a unidade nível 2 (microrregião);

$\gamma_n$ = coeficientes de inclinação para a unidade do nível 2 (microrregião);

$u_j$ = desvios das curvas em relação ao coeficiente de inclinação global.

Ou seja, a Equação 16 apresenta uma distribuição normal, com uma tendência central própria estimada pelo modelo e variância constante. Na Equação 15 observa-se que os erros aleatórios do primeiro nível e na Equação 18 (erros aleatórios do segundo nível) apresentam uma população normal, com média 0 e variância constante.

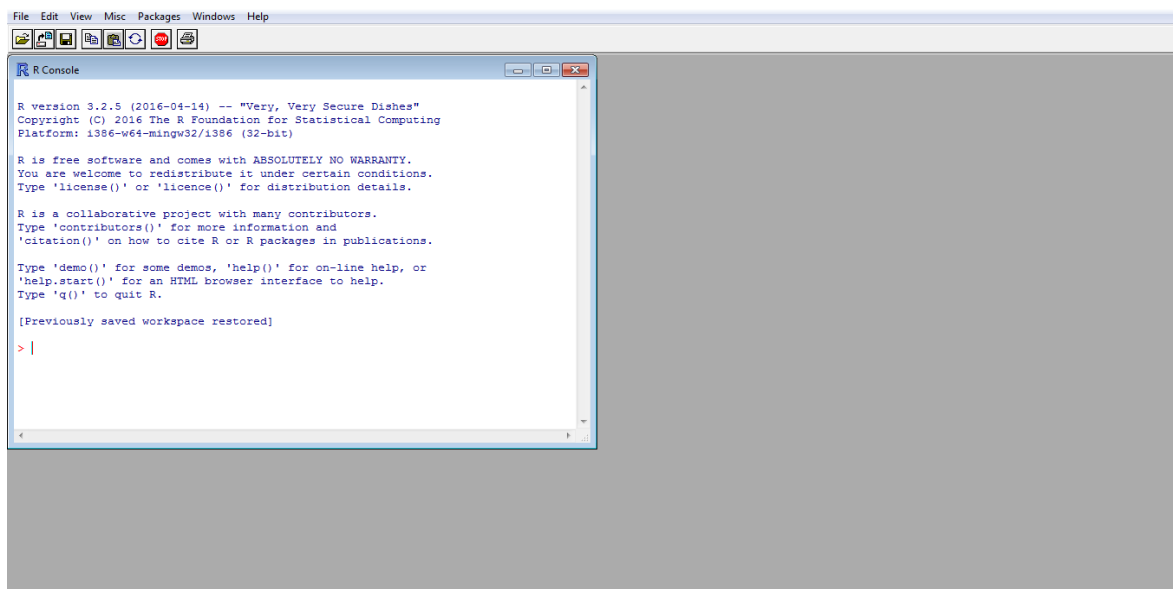
## 4 METODOLOGIA

### 4.1 SOFTWARE R (3.2.5)

O software R é um ambiente para computação gráfica e estatística, tratando-se de uma linguagem de programação especializada em computação com dados, apresentando algumas características marcantes, tais como a gratuidade e a sua disponibilidade para uma intensa gama de sistemas operacionais, sendo também altamente expansível ao uso de pacotes, apresentando-se com bibliotecas com funções específicas (SILVA; DINIZ; BORTOLUZZI, 2009).

O software R caracteriza-se como um programa de não elevada interatividade, apresentando-se no sistema operacional Windows, uma interface gráfica com barra de ferramentas no topo, permitindo realizar algumas tarefas comuns, tais como, abrir/gravar, cortar/colar textos, instalar pacotes, gerir programas, entre outros (OLIVEIRA, 2011) (Figura 7).

Figura 7 - Interface gráfica do software R (3.2.5)



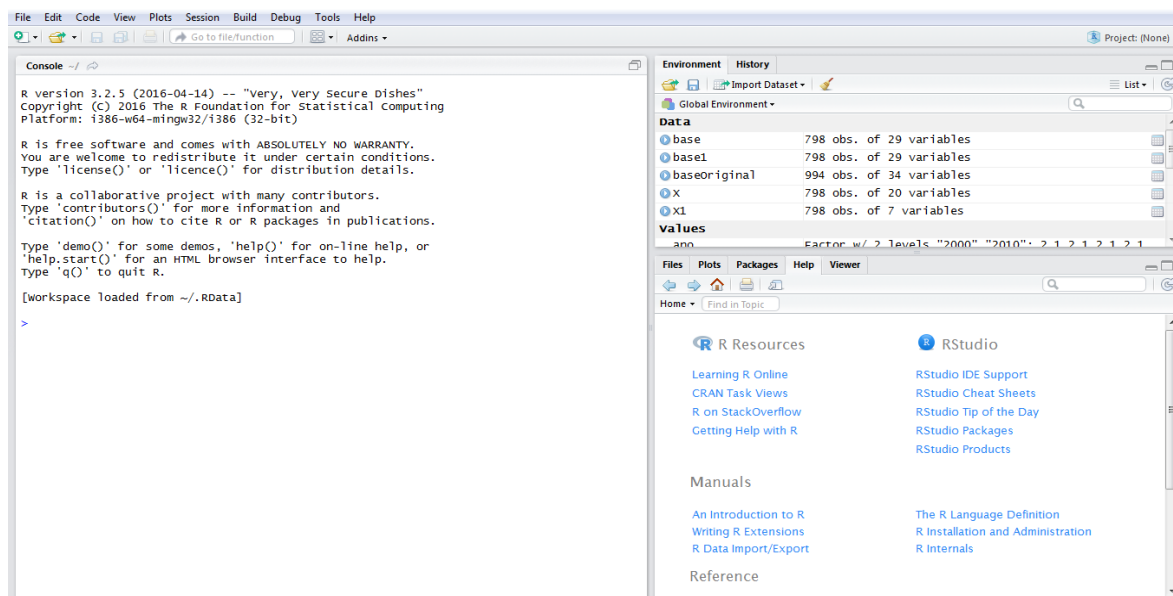
Fonte: Elaborado pelo autor, 2016.

O ambiente RStudio é uma interface gráfica que torna mais fácil e produtivo a utilização da linguagem R, sendo disponível para duas versões: RStudio Desktop (padrão e a

mais usual - utilizada neste estudo) e RStudio Server, o qual permite acessar um navegador web enquanto roda em um servidor GNU/Linux remoto (VERZANI, 2011).

A Figura 8 apresenta a interface gráfica do programa RStudio.

Figura 8 - Interface gráfica do programa RStudio



Fonte: Elaborado pelo autor, 2016.

Observa-se na Figura 8 a janela Console (esquerda). Nesta área que se desenvolvem os mais diversos algoritmos para implementação dos modelos requeridos ao estudo. Já na janela *Environment* apresenta-se um local em que há a possibilidade de salvar a base de dados e os resultados encontrados aos modelos propostos. E por fim, na janela direita inferior, encontram-se recursos como o menu *plots* (apresentação de gráficos elaborados aos modelos) e *Packages*, indicando os mais diversos tipos de pacotes que podem ser instalados no programa.

## 4.2 COLETA DE DADOS NOS SISTEMAS DE INFORMAÇÃO

Para o presente estudo, utilizou-se os sistemas de informação apresentados no Quadro 6, sendo estes utilizados para o cálculo e organização de indicadores de saúde ambiental.

Quadro 6 - Sistemas de informação estudados

<b>SISTEMAS DE INFORMAÇÃO</b>	<b>SIGLA</b>	<b>ENDEREÇO</b>
Departamento de Informática do SUS	DATASUS	<a href="http://www.datasus.gov.br">www.datasus.gov.br</a>
Atlas do Desenvolvimento Humano no Brasil 2013	ATLAS BRASIL	<a href="http://www.atlasbrasil.org.br">www.atlasbrasil.org.br</a>
Instituto de Pesquisa Econômica e Aplicada	IPEADATA	<a href="http://www.ipeadata.gov.br">http://www.ipeadata.gov.br</a>
Fundação de Economia e Estatística	FEEDADOS	<a href="http://feedados.fee.tche.br">http://feedados.fee.tche.br</a>

Fonte: Elaborado pelo autor, 2016.

É importante observar que a escolha dos sistemas de informação para obtenção de dados limitou-se em sites públicos.

#### 4.2.1 DATASUS

Utilizou-se o DATASUS para compor os indicadores descritos no Quadro 7.

Quadro 7 - Tipo de dado obtido no DATASUS e respectiva descrição  
(continua)

<b>DADO</b>	<b>DESCRIÇÃO DATASUS</b>
PIB per capita	Valor do PIB municipal per capita, calculado como sendo o PIB municipal do ano dividido pela população do mesmo ano. Os valores são apresentados em reais correntes, não sendo aplicado nenhum deflator ou fator de correção.
Ausência de rede de esgoto	Inexistência de um sistema onde fezes e urina são canalizadas para o sistema de esgoto (rede pública geral).
Famílias sem atendimento com rede de água	Famílias não servidas de água proveniente de uma rede geral de abastecimento, no domicílio ou no peridomicílio.
Famílias sem tratamento de água	Famílias que não possuem nenhum tipo de tratamento de água que é utilizada em seus domicílios.

(continuação)

<b>DADO</b>	<b>DESCRIÇÃO DATASUS</b>
Famílias sem coleta e com disposição inadequada de resíduo sólido	Famílias que não possuem coleta de resíduo sólido e o dispõem de forma inadequada (queimado/enterrado/disposto a céu aberto).
Morbidade por doenças infecciosas e parasitárias	Internações devido a doenças infecciosas e parasitárias.
Morbidade por enfermidades diarreicas e gastroenterite	Internações de pessoas devido a doenças diarreicas agudas, de causas infecciosas e parasitárias.
Mortalidade por doenças infecciosas e parasitárias	Óbitos devido a doenças infecciosas e parasitárias.
Mortalidade por enfermidades diarreicas e gastroenterite	Óbitos devido a doenças diarreicas agudas de causas infecciosas e parasitárias.
Número de internações hospitalares	Quantidade de internações pagas no período, não considerando as de prorrogação (longa permanência). Este é um valor aproximado das internações, pois as transferências e reinternações também são computadas, inclusive a dos crônicos e psiquiátricos que ultrapassaram o período máximo permitido.
Taxa de analfabetismo	Percentual de pessoas com 15 anos ou mais de idade que não sabem ler e escrever pelo menos um bilhete simples, no idioma que conhecem na população total residente da mesma faixa etária, em determinado espaço geográfico, no ano considerado.
Taxa de Desemprego de 16 anos ou mais	Proporção (%) da população residente economicamente ativa de 16 anos e mais que se encontra sem trabalho na semana de referência, em determinado espaço geográfico, no ano considerado.

Fonte: Informações extraídas do DATASUS ([www.datasus.gov.br](http://www.datasus.gov.br)). Elaborado pelo autor, 2016.

Nota: PIB per capita, Taxa de analfabetismo e Taxa de desemprego de 16 anos ou mais: dados disponibilizados no DATASUS com origem do IBGE.

Para o cálculo dos indicadores, referente aos dados apresentados no Quadro 7, foram utilizadas as metodologias apresentadas no Quadro 8.



Quadro 8 - Metodologia utilizada no cálculo dos indicadores de saúde ambiental para o DATASUS

INDICADOR	METODOLOGIA
PIB per capita; Taxa de analfabetismo; Taxa de desemprego de 16 anos ou mais;	Dados brutos
Ausência de rede de esgoto; Porcentagem de famílias sem atendimento com rede de água; Porcentagem de famílias sem tratamento de água; Porcentagem de famílias sem coleta e com disposição inadequada de resíduo sólido;	$INDICADOR = \frac{n^{\circ} \text{ de famílias com prevalência}}{n^{\circ} \text{ total de famílias}} \times 100$
Taxa de morbidade por doenças infecciosas parasitárias (geral e até 4 anos);	$IND = \frac{n^{\circ} \text{ de internações com prev.}}{n^{\circ} \text{ de residentes}} \times 1000$
Taxa de morbidade por enfermidades diarreicas e gastroenterite ;	$IND = \frac{n^{\circ} \text{ de internações com prev.}}{n^{\circ} \text{ total de residentes}} \times 10.000$
Taxa de internações hospitalares;	$IND = \frac{n^{\circ} \text{ de internações(geral)}}{n^{\circ} \text{ total de residentes}} \times 1000$
Taxa de mortalidade por doenças infecciosas e parasitárias;	$IND = \frac{n^{\circ} \text{ de óbitos com prevalência}}{n^{\circ} \text{ total de residentes}} \times 100000$
Taxa de mortalidade por enfermidades diarreicas e gastroenterite (geral e até 4 anos);	$IND = \frac{n^{\circ} \text{ de óbitos com prev.}}{n^{\circ} \text{ de residentes}} \times 100000$

Fonte: Elaborado pelo autor, 2016.

Nas equações descritas no Quadro 9, tem-se:

- $n^{\circ}$  de famílias com prevalência: número de famílias que apresentam o fator analisado no COREDE ou RF (por exemplo, número total de famílias que não possuem rede de distribuição de água no município);
- $n^{\circ}$  total de famílias: número total de famílias que apresentam as características do indicador estudado, nos COREDES ou RFs analisados (por exemplo, número total de famílias no município).

#### 4.2.2 ATLAS BRASIL

Utilizou-se do ATLAS BRASIL para compor os indicadores apresentados no Quadro 9 e respectiva definição de acordo com o próprio sistema de informação.

Quadro 9 - Tipo de dado obtido no ATLASBRASIL e respectiva descrição  
(continua)

<b>DADO</b>	<b>DESCRIÇÃO ATLAS BRASIL</b>
Expectativa de vida ao nascer	Expectativa de anos de vida de uma pessoa nascida no ano de referência supondo que as taxas de mortalidade por idade estimada para anos anteriores se mantivessem constantes nos anos posteriores.
Índice de Desenvolvimento Humano	Índice que avalia o grau de desenvolvimento econômico e a qualidade de vida oferecida a população local.
Pessoas vulneráveis à pobreza	Proporção dos indivíduos com renda domiciliar per capita igual ou inferior a R\$ 255,00 mensais, em reais de agosto de 2010, equivalente a ½ salário mínimo nessa data. O universo de indivíduos é limitado àqueles que vivem em domicílios particulares permanentes.
Pessoas acima de 18 anos desocupadas	Corresponde ao percentual da população economicamente ativa (PEA) nessa faixa etária que estava desocupada, ou seja, que não estava ocupada na semana anterior à data dessa pesquisa.
Ausência de coleta de resíduo sólido na área urbana	Razão entre a população que vive em domicílios sem coleta de lixo e a população total residente em domicílios particulares permanentes, multiplicada por 100. Estão incluídas as situações em que a coleta de lixo é realizada diretamente por empresa pública ou privada, e quando o lixo é depositado em caçamba, tanque ou depósito fora do domicílio, para posterior coleta por prestadora do serviço. São considerados apenas os domicílios particulares permanentes localizados em área urbana.

(continuação)

<b>DADO</b>	<b>DESCRIÇÃO ATLAS BRASIL</b>
Pessoas com banheiro e água encanada	Razão entre a população que vive em domicílios particulares permanentes com água encanada em pelo menos um de seus cômodos e com banheiro exclusivo e a população total residente em domicílios particulares permanentes, multiplicadas por 100. A água pode ser proveniente de rede geral, de poço, de nascente ou de reservatório abastecido por água das chuvas ou carro-pipa. Banheiro exclusivo é definido como cômodo que dispõe de chuveiro ou banheira e aparelho sanitário.
Pessoas com abastecimento de água e esgotamento sanitário inadequado	Razão entre as pessoas que vivem em domicílios cujo abastecimento de água não provém de rede geral e cujo esgotamento sanitário não é realizado por rede coletora de esgoto ou fossa séptica, e a população total residente em domicílios particulares permanentes, multiplicada por 100. São considerados apenas os domicílios particulares permanentes.

Fonte: Informações extraídas do ATLAS 2013 (<http://www.atlasbrasil.org.br/>). Elaborado pelo autor, 2016.

Os dados relacionados no Quadro 9 apresentam-se na forma de porcentagem, não necessitando cálculos ou manipulações matemáticas para obtenção dos indicadores.

#### 4.2.3 ATLAS DA VULNERABILIDADE SOCIAL

É um sistema de informação capaz de evidenciar os problemas sociais nos diferentes municípios dos Estados federativos brasileiros.

No Quadro 10 são apresentados os indicadores coletados neste sistema de informação.

Quadro 10 - Tipo de dado do Atlas da Vulnerabilidade Social e respectiva descrição  
(continua)

<b>TIPO DADO</b>	<b>DESCRIÇÃO ATLAS DA VULNERABILIDADE SOCIAL</b>
Índice de Vulnerabilidade Social (IVS)	Complementa o Índice de Desenvolvimento Humano Municipal (IDHM). O IVS traz dezesseis indicadores estruturados em três dimensões, a saber, infraestrutura urbana, capital humano e renda e trabalho, permitindo um

(continuação)

<b>TIPO DADO</b>	<b>DESCRIÇÃO ATLAS DA VULNERABILIDADE SOCIAL</b>
	mapeamento singular da exclusão e da vulnerabilidade social para os 5.565 municípios brasileiros.
Porcentagem de crianças em domicílios em que ninguém tem ensino fundamental completo	Razão entre o número de crianças de até 14 anos que vivem em domicílios em que nenhum dos moradores tem o ensino fundamental completo, e a população total nesta faixa etária residente em domicílios particulares permanentes (multiplicada por 100).
Porcentagem de mulheres de 10 a 17 anos que tiveram filhos	Razão entre o número de mulheres de 10 a 17 anos de idade que tiveram filhos, e o total de mulheres nesta faixa etária (multiplicada por 100).

Fonte: Informações extraídas de [http://ivs.ipea.gov.br/ivs/data/rawData/publicacao\\_atlas\\_ivs.pdf](http://ivs.ipea.gov.br/ivs/data/rawData/publicacao_atlas_ivs.pdf). Elaborado pelo autor, 2016.

Todos os dados constantes no Quadro 10 apresentam-se na forma de porcentagem, não necessitando cálculos ou manipulações matemáticas para obtenção dos indicadores.

#### 4.2.4 IPEADATA

Com relação ao sistema de informação IPEADATA, utilizou-se os dados apresentados no Quadro 11.

Quadro 11 - Tipo de dado do IPEADATA e respectiva descrição

<b>DADO</b>	<b>DESCRIÇÃO ATLAS BRASIL</b>
Médicos residentes por 1.000 hab.	Proporção de médicos por mil habitantes.
Média de Anos de Estudo	Razão entre o somatório do número de anos de estudo completados por pessoas com 25 anos ou mais e o número de pessoas compreendidas nesta faixa etária.

Fonte: Informações extraídas de IPEADATA (<http://www.ipeadata.gov.br/>). Elaborado pelo autor, 2016.

Para os dados observados no Quadro 11 não houve a necessidade de cálculos ou manipulações matemáticas para obtenção dos indicadores estudados.

#### 4.2.5 FEEDADOS

Para o sistema de informação FEEDADOS, utilizou-se os dados apresentados no Quadro 12.

Quadro 12 - Tipo de dado da FEEDADOS e respectiva descrição

<b>DADO</b>	<b>DESCRIÇÃO ATLAS BRASIL</b>
Número de leitos hospitalares	Número de leitos hospitalares encontrados no COREDE
Hospitais	Número de Hospitais encontrados no COREDE

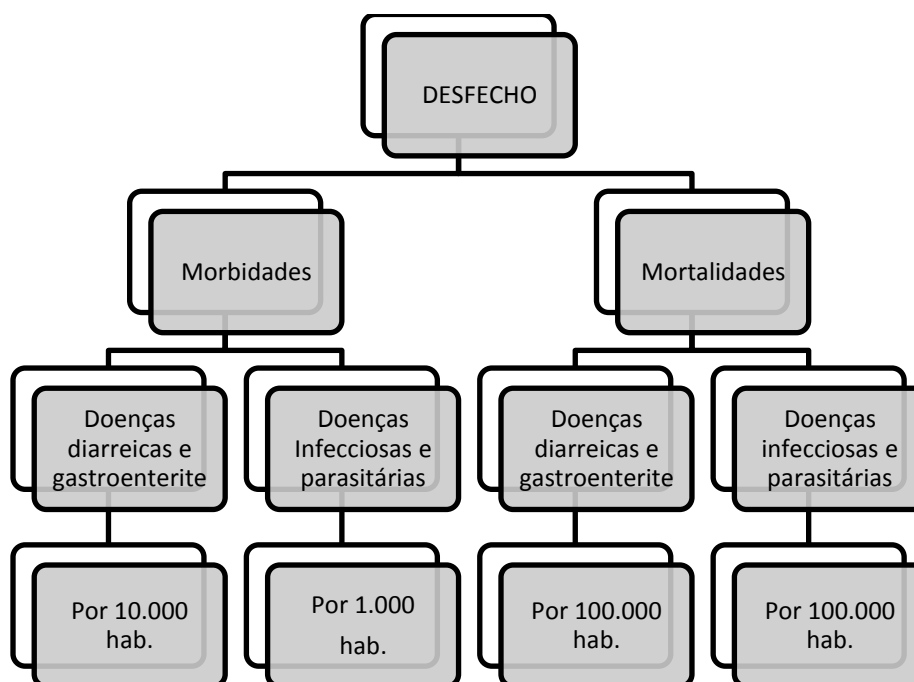
Fonte: Informações extraídas de FEEDADOS (<http://feedados.fee.tche.br/>). Elaborado pelo autor, 2016.

O indicador “Leitos hospitalares por 1.000 habitantes” foi definido pela divisão do número de leitos hospitalares por 1.000. O indicador “Hospitais” foi considerado como o dado bruto disponível na FEEDADOS.

#### 4.3 DESFECHOS

Na Figura 9 são apresentados os desfechos utilizados nos modelos de regressão linear múltipla e modelos multiníveis.

Figura 9 – Desfechos dos modelos de regressão linear múltipla e modelos multiníveis



Fonte: Elaborado pelo autor, 2016.

Buscou-se desenvolver modelos estatísticos para descrever o comportamento dos coeficientes de morbidades e mortalidades para as doenças diarreicas e gastroenterite, como também, para as doenças infecciosas e parasitárias das microrregiões do Estado do RS.

#### 4.4 ESTRATÉGIAS DE MODELAGEM

##### 4.4.1 Modelos de regressão linear múltipla

###### 4.4.1.1 ETAPA 1: Elaboração dos modelos

Durante a elaboração dos modelos estatísticos utilizou-se a estratégia para seleção dos preditores significativos, conhecida como *stepwise backward selection* a qual elimina um preditor por vez, considerando todos os preditores na base de dados. Vale destacar que o software R (3.2.5) utiliza o aumento do AIC como critério para manutenção das variáveis.

A técnica *backward* incorpora, inicialmente, todas as variáveis no modelo e depois, por etapas, cada uma pode ser ou não eliminada (FERREIRA, 2012).

Utilizou-se do critério de adição ou remoção das variáveis, baseado na estatística de Fischer (Teste F), comparando-se desta forma, modelos com ou sem os preditores disponíveis na base de dados.

Observa-se que foram coletados os dados para todo o período disponível em cada sistema de informação. Porém, em função dos preditores não estarem disponíveis para todos os anos, o estudo baseou-se nos anos de 2000 e 2010.

O processo de desenvolvimento dos modelos de regressão linear múltipla se desenvolveu de maneira simples, uma vez que a própria estratégia *backward* realiza a escolha do melhor ajuste ao desfecho analisado. Exemplifica-se a técnica *stepwise backward selection* na Equação 17.

$$> mstep = step(lm(\text{`Coeficiente morbidade por doenças infecciosas e parasitárias' ~ .}, data = na.exclude(base2)), direction = 'backward') \quad \text{Eq. 17}$$

Através de um procedimento automatizado, o algoritmo da Equação 17 apresenta-nos os preditores com nível de significância menor que 0,05 ( $p < 0,05$ ) para determinado modelo proposto ao coeficiente em estudo. Por fim, utilizou-se do comando *summary* no software R (3.2.5) para visualizar os respectivos coeficientes e preditores significativos ao modelo elaborado. A Equação 18 exemplifica o caso.

$$> summary(mstep) \quad \text{Eq. 18}$$

#### 4.5.1.2 ETAPA 2: Definição do preditor mais significativo

Após a elaboração dos modelos finais relacionados aos desfechos, buscou-se verificar o preditor com maior nível de significância. Para isto, utilizou-se do comando ‘Anova’, exemplificado na Equação 19.

$$> anova(mstep) \quad \text{Eq. 19}$$

Após, buscou-se verificar um comportamento gráfico entre o preditor com maior nível de significância e o desfecho analisado, como sendo apresentado na Equação 20.

$$M <- lm(base\$y \sim x) \quad \text{Eq. 20}$$

Ou seja:

$$M3 <- lm(base3\$ \text{`Coeficiente mortalidade por doenças infecciosas e} \quad \text{Eq. 21}$$

*parasitárias`~base3\$`Porcentagem de pessoas acima de 18 anos  
desocupadas`)*

Por fim, elaborou-se o gráfico que relaciona o desfecho com o preditor de maior nível de significância (Equação 22).

$$Plot(M) \quad \text{Eq. 22}$$

Uma observação importante é que o estudo com modelos de regressão linear múltipla não contemplou os preditores do Atlas da Vulnerabilidade Social, compreendido pelos indicadores “Porcentagem de mulheres de 10 a 17 anos que tiveram filhos”, “Porcentagem de crianças em domicílios em que ninguém tem ensino fundamental completo”, “Taxa de envelhecimento” e “Índice de vulnerabilidade social”.

#### 4.4.2 Estratégias de modelagem multinível

Utilizou-se do software R (3.2.5) para a elaboração dos modelos multiníveis que descrevessem os desfechos analisados. A realização da modelagem se desenvolveu por etapas, as quais estão descritas a seguir.

##### 4.4.2.1 ETAPA 1: Remoção de *missing* superior a 20%

Primeiramente, as variáveis com um número superior a 20% de dados faltantes (*missing*) foram removidas do banco de dados. O ponto de corte (20%) foi definido pelo autor. Esta etapa foi necessária, pois a base de dados disponível para a construção dos modelos possuía informações não disponíveis para diversos anos referente aos municípios que compõem as microrregiões do Estado do RS.

O algoritmo desenvolvido no software R (3.2.5) que calcula os preditores com grande número de dados faltantes é exemplificado nas Equações 23 e 24.

$$> \text{baseOriginal} <- \text{base} \quad \text{Eq.23}$$

$$> \text{base} <- \text{baseOriginal}[, \text{apply}(\text{baseOriginal}, 2, \text{function}(x) \text{mean}(\text{is.na}(x))) < 0.2] \quad \text{Eq. 24}$$

Na sequência, foram excluídos todos os casos restantes com *missing* superior a 20%. Exemplifica-se o caso na Equação 25.



$$>base <- na.exclude(base) \quad \text{Eq. 25}$$

Uma observação importante é que grande parte dos indicadores com *missing* superior a 20% compuseram-se por indicadores calculados através de dados fornecidos pela Estratégia da Saúde da Família (ESF), disponíveis no DATASUS, do Ministério da Saúde.

#### 4.4.2.2 ETAPA 2: Extração da Região e Ano

Optou-se pela elaboração de dois comandos que realizassem a extração da base de dados (Região e Ano), facilitando posteriormente a organização do algoritmo. A Equação 26 exemplifica o caso.

$$\begin{aligned} >regiao <- base\$regiao \\ >ano <- as.factor(base\$ano) \end{aligned} \quad \text{Eq. 26}$$

#### 4.4.2.3 ETAPA 3: Extração da variável desfecho de interesse

Elaborou-se um algoritmo que extraísse a variável desfecho de interesse, devendo esta, ser alterada de acordo com o modelo a ser ajustado, como exemplificado na Equação 27.

$$>y <- base\$`Coeficiente mortalidade por doenças infecciosas e parasitárias` \quad \text{Eq. 27}$$

Observa-se que este algoritmo não considera a variação da escala original ou logarítmica do melhor modelo ajustado, o que se faz necessário uma análise manual de ambos os modelos, de acordo com seus respectivos gráficos de resíduos.

Conforme necessidade da utilização da escala logarítmica no desfecho, utiliza-se como exemplo a Equação 28.

$$>y <- \log(base\$`E3- Taxa mortalidade por doenças infecciosas e parasitárias`+0.1) \quad \text{Eq. 28}$$

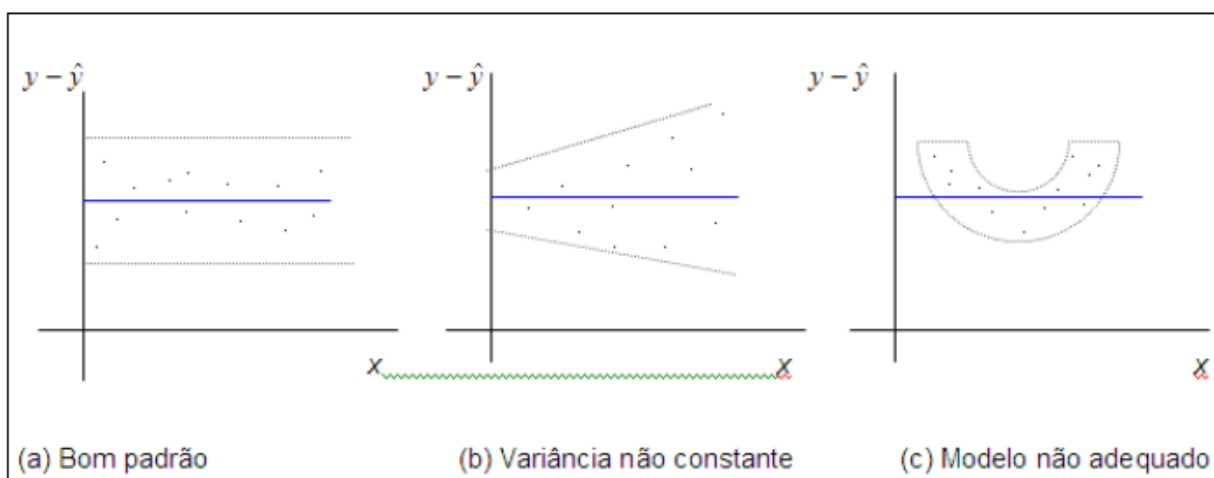
Observa-se a utilização de uma pequena constante aliada ao desfecho de interesse (0,1), uma vez que para valores com coeficiente zero, a função é indeterminada.

#### 4.4.2.4 ETAPA 4: Verificação do gráfico de resíduos

Nesta etapa foi realizada a análise da melhor disposição dos pontos na reta do gráfico de resíduos.

De acordo com Battisti (2013), o modelo com ajuste ideal das variáveis preditoras apresenta-se no gráfico de resíduos ilustrado na Figura 10 a).

Figura 10 - Gráficos para análise de resíduos em regressão



Fonte: BATTISTI (2013).

Desta forma optou-se pelo melhor modelo que descreveu o coeficiente em análise (escala original ou logarítmica).

#### 4.4.2.5 ETAPA 5: Retirada das variáveis que não podem ser utilizadas como preditoras

Nesta etapa, desenvolveu-se um comando para a exclusão de todas as variáveis que estão impossibilitadas de serem incluídas como variáveis preditores aos modelos propostos (QUADRO 13).

Quadro 13 - Variáveis não utilizadas como preditoras nos modelos multiníveis

<b>VARIÁVEL</b>	<b>COLUNA CORRESPONDENTE SOFTWARE R</b>
Município	1
Ano	2
Região	3
Coeficiente de morbidade por doenças infecciosas e parasitárias até 4 anos	10
Coeficiente de morbidade por doenças infecciosas e parasitárias	11
Coeficiente de mortalidade por doenças infecciosas e parasitárias	12
Coeficiente de morbidade por doenças diarreicas e gastroenterite	24
Coeficiente de mortalidade por doenças diarreicas e gastroenterite	25
Coeficiente de mortalidade por doenças diarreicas e gastroenterite até 4 anos	26

Fonte: Elaborado pelo autor, 2016.

Como observado no Quadro 13, as variáveis que não podem ser consideradas como preditoras são os coeficientes (desfechos), Ano, Região e Município. Este procedimento faz-se necessário uma vez que durante a técnica utilizada para elaboração dos modelos estatísticos, o algoritmo não apresente a possibilidade de considerar estas variáveis como preditoras aos modelos propostos.

#### 4.4.2.6 ETAPA 6 (alternativa): Exclusão de zeros na base de dados

Esta etapa é necessária quando verifica-se uma falta de ajuste dos pontos nos gráficos de resíduos, tanto em estala normal como logarítmica.

Optou-se pela criação de um vetor booleano, representado por operadores que retornam na forma 'TRUE' ou 'FALSE', indicando quais casos da variável desfecho é igual à zero.

O primeiro passo é definir o nome da variável e em seguida atribuir a ela o resultado da operação que se deseja (para o caso do estudo, ser diferente de zero), como apresentado na Equação 29.

$$>sel = base\$`Coeficiente mortalidade por doenças infecciosas e parasitárias`!= 0 \quad \text{Eq. 29}$$

Utilizou-se essa variável para filtrar as demais, tais como: desfecho, ano e região (Equações 30, 31, 32 e 33).

$$>y = y[sel] \quad \text{Eq. 30}$$

$$>ano = ano[sel] \quad \text{Eq. 31}$$

$$>Região = Região[sel] \quad \text{Eq. 32}$$

$$>X = X[sel, ] \quad \text{Eq. 33}$$

#### 4.4.2.7 ETAPA 7: Função para a busca do melhor modelo

Nesta etapa considera-se o desenvolvimento do algoritmo para a busca do melhor modelo ao desfecho analisado, utilizando como função objetiva o AIC (*Akaike Information Criterion*) e a estratégia *Simulated Annealing* (método de otimização) para explorar os modelos possíveis.

Aplicam-se diferentes metodologias para a seleção de modelos, tais como  $C_p$  de Mallows, Regressão Stepwise, Critério de Informação Akaike (AIC), Critério de Informação Bayesiano (BIC), dentre outros (EMILIANO, 2009).

Os critérios de informação são largamente usados como forma de auxiliar na escolha da ordem adequada do modelo. Dentre os critérios de informação, o mais conhecido e ainda largamente utilizado, é o Critério de Informação de Akaike (AIC – *Akaike Information Criterion*), em que se baseia na divergência Kullback-Leibler, que é uma medida de quanta informação é “perdida” ao tentar representar um conjunto  $T$  de medidas utilizando uma base conhecida  $L$ . Porém, como não se conhece o modelo que gera  $T$ , não é possível comparar o modelo obtido ao modelo “ideal”  $T$ . Entretanto, Akaike formulou um meio de fazer um cálculo aproximado da divergência de Kullback-Leibler com relativa simplicidade, que provavelmente, seja esta a razão de sua grande aceitação (SOBRAL; BARRETO, 2011).

Nos critérios de informação AIC e BIC, para cada modelo gera-se um valor de AIC ou BIC, sendo que o melhor modelo é representado por seu respectivo menor valor (EMILIANO, 2009).

Já o *Simulated Annealing* é o método probabilístico proposto em Kirkpatrick, Gelett e Vecchi (1983) e Cerny (1985) para encontrar o mínimo global de uma função de custo que pode possuir vários mínimos locais. Ele funciona emulando o processo físico pelo qual um sólido é lentamente resfriado de modo que quando sua estrutura é finalmente "congelada", isso acontece em uma configuração de energia mínima (BERTSIMAS; TSITSIKLIS, 1993).

Neste estudo, temos uma função objetiva (AIC) que nos dá a “energia” do modelo. Queremos então encontrar um subconjunto (estados) de preditores que apresentem o menor valor de AIC. Pela não facilidade de exploração de todos os estados possíveis, a solução se dá por anelamento simulado, realizando-se da seguinte forma:

1. Inicia em um estado “x” (determinado subconjunto aleatório de preditores);
2. Verifica o AIC deste modelo;
3. Executa uma modificação aleatória:
  - 3.1. Adiciona um preditor (log ou normal);
  - 3.2. Exclui um preditor;
4. Verifica o AIC do novo modelo;
5. Se o AIC é menor: aceitar o novo modelo; Se AIC for maior (Equação 34):

$$\Pr(M_{atual} \rightarrow M_{proposto}) = e^{\left[ \frac{AIC_{modelo_{novo}} - AIC_{modelo_{antigo}}}{T} \right]} \quad \text{Eq. 34}$$

Onde:

T= temperatura.

Quando T é alto, o algoritmo se move mais livremente entre os diferentes modelos, mesmo quando o modelo é um pouco pior. Durante a execução do algoritmo, a temperatura é reduzida lentamente (por isso a analogia com a metalurgia), de forma que fica cada vez mais difícil de aceitar um modelo que piore o AIC. Dessa forma, consegue-se explorar o espaço dos estados (os diferentes modelos) com amplitude no início, e, à medida em que este vai “resfriando”, aproximamo-nos do mínimo global.

O algoritmo apresentado na Equação 35 é de autoria de Erikson Kazubowski.

$$\begin{aligned} &>searchM <- function(y, ano, regioao, X, logC=0.1, \\ &\quad initTemp=8, stages=15, rate=0.8, \\ &\quad verbose=TRUE) \{ \\ &\quad \quad iters = rep(NA, stages) \\ &\quad \quad for (i in 1:stages) \{ \end{aligned} \quad \text{Eq. 35}$$

```

    iters[i] = round(20*1.2^i)
  }

  p <- dim(X)[2]
  vars <- array(NA, c(sum(iters), p))
  AICs <- rep(NA, sum(iters))
  whichX <- sample(0:2, p, replace=T, prob=c(0.5, 0.25, 0.25))

  X1 <- X

  if (2 %in% whichX){
    X1[,whichX == 2] <- log(X1[,whichX == 2]+logC)
  }
  names(X1)[whichX==2] <- paste('log', names(X1)[whichX==2])
  X1 <- scale(as.matrix(X1[,whichX > 0]))
  currM <- lmer(y~ano+X1+(1+ano|regiao), REML=FALSE,
    control=lmerControl(optimizer='Nelder_Mead'))
  minM <- currM
  currAIC <- AIC(currM)
  vars[1, ] <- whichX
  AICs[1] <- currAIC

  temp = initTemp
  cumIters = c(2, cumsum(iters))

  for (m in 1:(stages)){

    temp = temp * rate

    for (i in cumIters[m]:cumIters[m+1]){
      changeVar <- sample(1:p, 1)
      whichXprop <- vars[i-1, ]
      whichXprop[changeVar] <- sample((0:2)[- (whichXprop[changeVar]+1)],
        1)
      X1 <- X

      if (2 %in% whichXprop){
        X1[, whichXprop==2] <- log(X1[, whichXprop==2]+logC)
      }
      names(X1)[whichXprop==2] <- paste('log', names(X1)[whichXprop==2])
      X1 <- scale(as.matrix(X1[, whichXprop>0]))
      tryCatch({
        lm1 <- lmer(y~ano+X1+(1+ano|regiao), REML=FALSE,
          control=lmerControl(optimizer='Nelder_Mead'))
      }, warning=function(w) {
        cat('Warning!\n')
      }, error=function(e) {
        cat('whichX:', whichXprop, '\n\n')
      })
    }
  }

```

```

        print('Error!\n')
        cat('whichX: ', whichXprop, '\n\n')
    })

    propAIC <- AIC(lm1)
    if(verbose){
        cat('Temp: ', temp, '\n')
        cat('Iter: ', i, '\n')
        cat('Total: ', round(i/max(cumIters)*100), '%\n')
        cat('propAIC: ', propAIC, ', currAIC: ', currAIC, '\n')
        cat('deltaAIC: ', (currAIC-propAIC), '\n')
        cat('Acc. prob: ', min(1, exp((currAIC - propAIC)/temp)), '\n\n')

    }

    if (propAIC < min(AICs, na.rm=T)) {
        minM <- lm1
    }

    if ((propAIC<currAIC)){
        vars[i, ] <- whichXprop
        currAIC <- propAIC
        AICs[i] <- propAIC
        currM <- lm1
    } else {
        if (exp((currAIC - propAIC)/temp) > runif(1)){
            vars[i, ] <- whichXprop
            currAIC <- propAIC
            AICs[i] <- propAIC
            currM <- lm1
        } else {
            vars[i, ] <- vars[i-1, ]
            AICs[i] <- currAIC
        }
    }
}

list(M=currM, minM=minM, vars=vars, AICs=AICs)
}

```

Ou seja, o algoritmo apresentado na Equação 35, realiza o cálculo de acordo com a variável desfecho (y) escolhida, a qual é alterada de acordo com a Etapa 3, bem como as variáveis ano, região e X, não necessitam serem alteradas nesta função.

Observa-se também uma pequena constante, representada por 'logC', com o valor fixo de 0,1, apresentando com isto, valor determinado em todas as situações.

#### 4.4.2.8 ETAPA 8: Resultado final dos AICs

Elaborou-se uma variável que pudesse apresentar o resultado final das diferentes equações elaboradas no software R (3.2.5), verificando com isto, o menor AIC presente nas equações ajustadas (Equação 36).

$$>out <- searchM(y, ano, regioao, X) \quad \text{Eq. 36}$$

#### 4.4.2.9 ETAPA 9: Busca pelo melhor modelo

A cada resultado visualizado no ambiente de trabalho do RStudio, há a opção de armazenamento dos dados no espaço *Environment*. Para isto, optou-se por comandos com nomes de acordo com as iniciais do desfecho em estudo, como apresentado na Equação 37.

$$>morbigastr01=out \quad \text{Eq. 37}$$

Assim, altera-se apenas o número final da variável estudada.

De maneira geral, optou-se por rodar cerca de 20 vezes o algoritmo para o mesmo modelo do desfecho analisado.

#### 4.4.2.10 ETAPA 10: Modelo final

Após sucessivas aplicações do algoritmo para o desfecho analisado, verifica-se o modelo com menor valor de AIC, uma vez que em modelos multiníveis, uma medida do "melhor" é dada pelo menor AIC. Para isto, utiliza-se do comando *summary* para cada variável salva aos modelos encontrados anteriormente (Equação 38).

$$>summary(morbigastr01\$minM) \quad \text{Eq. 38}$$

Desta forma, optando-se pelo *minM*, os valores dos preditores se encontram com dimensão *escore-z*, facilitando a análise dos modelos finais.



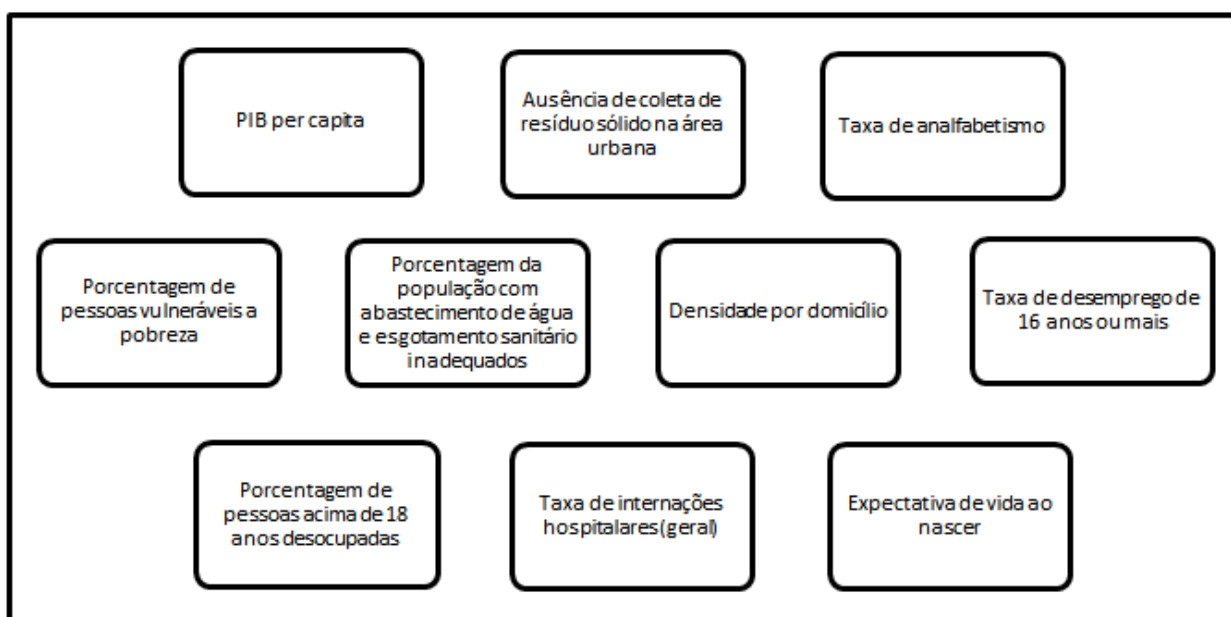
## 5 RESULTADOS E DISCUSSÕES

Primeiramente, serão apresentados os resultados do ajuste de modelos de regressão considerando a modelagem tradicional e na sequência serão apresentados os resultados do ajuste de modelos multiníveis. Para cada metodologia foram ajustadas quatro equações considerando diferentes desfechos.

### 5.1 MODELOS DE REGRESSÃO LINEAR MÚLTIPLA - MODELAGEM TRADICIONAL

Os preditores, selecionados na ETAPA 1 da modelagem, com menos de 20% de dados faltantes são apresentados na Figura 11.

Figura 11 - Preditores com menos de 20% de dados faltantes para a construção dos modelos de regressão linear múltipla



Fonte: Elaborado pelo autor, 2016.

### 5.1.1 Modelo para o desfecho “Coeficiente de morbidade por doenças infecciosas e parasitárias” por 1.000 habitantes (modelo 1)

As variáveis preditores do modelo 1 final, obtidos pela técnica *stepwise backward selection* estão apresentada na Tabela 1.

Tabela 1 - Regressão linear multivariada para o modelo 1 final dos COREDEs do Estado do RS nos anos 2000 e 2010

R <sup>2</sup>	Intercepto (coeficiente ± erro padrão)	Variáveis preditoras	Coeficiente $\beta$	Erro-padrão
0,2154	579,95±97,83	Ano	-0,29	0,05
		Microrregiões do RS	-	-
		Porcentagem da população com banheiro e água encanada	-0,056	0,027
		Porcentagem da população com abastecimento de água e esgotamento sanitário inadequados	-0,12	0,05
		IDHM	11,91	4,55

Fonte: Elaborado pelo autor, 2016. Nível de significância < 0,05.

Analisando-se os dados apresentados na Tabela 1, verifica-se que a variável preditora “Ano” (variável binária), indica que de 2000 para 2010 houve uma redução de 0,29 casos a cada 1.000 habitantes no modelo 1 final.

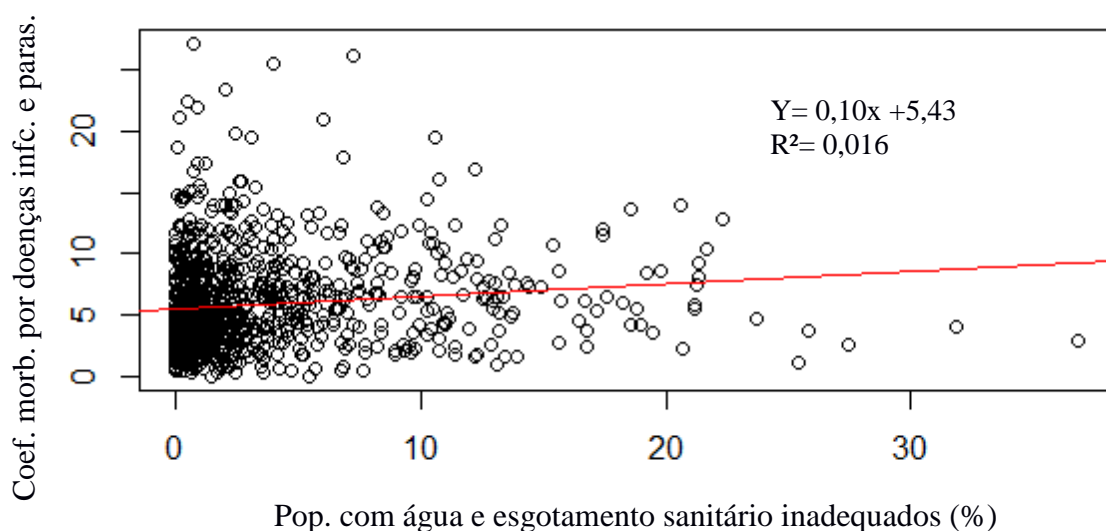
Também, pode ser observado na Tabela 1, que para cada ponto percentual a mais de “População com banheiro e água encanada” e de “População com abastecimento de água e esgotamento sanitário inadequados”, houve um decréscimo de 0,056 casos e 0,12 casos a cada 1.000 habitantes no modelo 1 final, respectivamente.

Também, a variável categórica “Microrregiões” apresenta-se importante ao modelo, evidenciando com isto, diferenças entre os COREDEs quando relacionada ao modelo 1 final.

Por fim, a variável preditora “IDHM”, no ano de 2010, apresentou uma contribuição de 11,91 casos no modelo 1 final a cada 1.000 habitantes das microrregiões do RS.

A Figura 12 apresenta a relação entre o “Coeficiente de morbidade por doenças infecciosas e parasitárias” (2000 e 2010) e a “Porcentagem da população com abastecimento de água e esgotamento sanitários inadequados” nas microrregiões do RS (COREDEs) (2000 e 2010). A escolha desta variável deu-se pela Análise de Variância, a qual indicou-a como a variável preditora mais relevante (mais significativa) para o modelo.

Figura 12 - Regressão linear simples entre “Coeficiente de morbidade por doenças infecciosas e parasitárias” e “Porcentagem da população com abastecimento de água e esgotamento sanitários inadequados”, COREDEs do RS (2000 e 2010)



Fonte: Elaborado pelo autor, 2016.

Observa-se na Figura 12, uma relação positiva entre a “População com abastecimento de água e esgotamento sanitário inadequados” e o desfecho “Coeficiente de morbidade por doenças infecciosas e parasitárias”, em que há uma contribuição de 0,10 vezes no “Coeficiente de morbidade por doenças infecciosas e parasitárias” a cada unidade da “População com abastecimento de água e esgotamento sanitário inadequados”. Porém, a reta ajustada apresentou um coeficiente de determinação de 1,6%, indicando um péssimo ajuste dos pontos.

### 5.1.2 Modelo para o desfecho “Coeficiente de mortalidade por doenças infecciosas e parasitárias” por 100.000 habitantes (modelo 2)

Através da técnica *stepwise backward selection*, o modelo 2 final foi composto pelas variáveis preditoras apresentadas na Tabela 2.

Tabela 2 - Regressão linear multivariada para o modelo 2 final dos COREDEs do Estado do RS nos anos 2000 e 2010

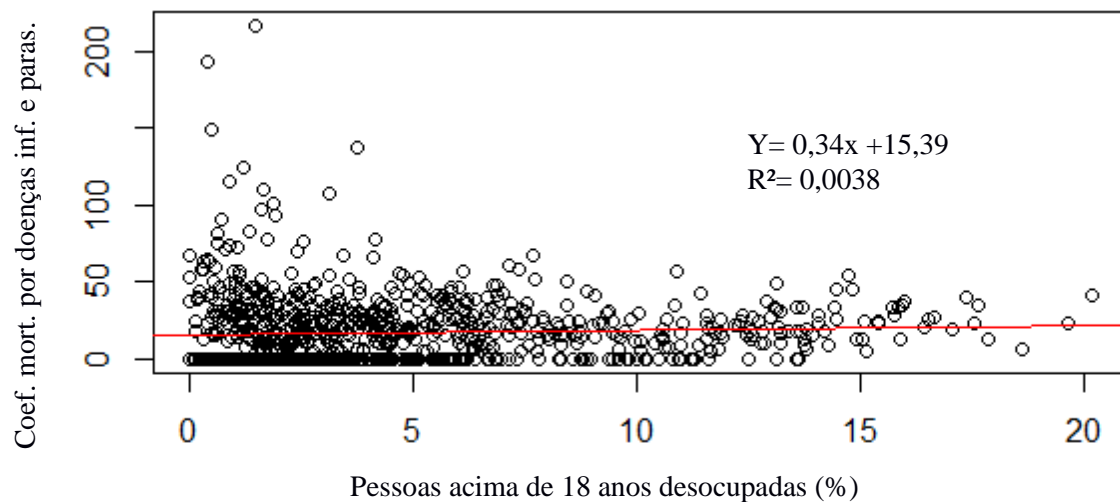
R <sup>2</sup>	Intercepto (coeficiente ± erro padrão)	Variáveis preditoras	Coeficiente $\beta$	Erro-padrão
0,054	-1,706392	Ano	0,84	0,2
		Porcentagem de pessoas acima de 18 anos desocupadas	0,55	0,2
		Densidade por Domicílio	0,38	0,1
		Expectativa de Vida ao Nascer	1,35	0,5

Fonte: Elaborado pelo autor, 2016. Nível de significância < 0,05.

Na Tabela 2 observa-se que a variável preditora “Ano”, apresentou uma contribuição de 0,84 casos no modelo 2 final a cada 100.000 pessoas residentes nas microrregiões do Estado do RS, isto é, no ano de 2010 teve um aumento de 0,84 casos por 100.000 habitantes em relação a 2000.

A Figura 13 apresenta a relação entre a “Porcentagem de pessoas acima de 18 anos desocupadas” (2000 e 2010) e o “Coeficiente de mortalidade por doenças infecciosas e parasitárias” (2000 e 2010).

Figura 13 - Regressão linear simples entre “Coeficiente de mortalidade por doenças infecciosas e parasitárias” e “Porcentagem de pessoas acima de 18 anos desocupadas”, COREDEs do RS (2000 e 2010)



Fonte: Elaborado pelo autor, 2016.

Ao analisar a Figura 13, observa-se que a “Porcentagem de pessoas acima de 18 anos desocupadas”, contribui em 0,34 vezes o “Coeficiente de mortalidade por doenças infecciosas e parasitárias”. Contudo, a equação apresentou um coeficiente de determinação 0,38%, indicando um péssimo ajuste da reta aos pontos.

### 5.1.3 Modelo para o desfecho “Coeficiente de mortalidade por doenças diarreicas e gastroenterite” por 100.000 habitantes (modelo 3)

O modelo 3 final é composto pelas variáveis preditoras apresentadas na Tabela 3, através da técnica *stepwise backward selection*.

Tabela 3 - Regressão linear multivariada para o modelo 3 final dos COREDEs do Estado do RS nos anos 2000 e 2010

(continua)

R <sup>2</sup>	Intercepto (coeficiente ± erro padrão)	Variáveis preditoras	Coeficiente $\beta$	Erro-padrão
0,01	367,27±123,4	Ano	0,18	0,06

(continuação)

<b>R<sup>2</sup></b>	<b>Intercepto (coeficiente ± erro padrão)</b>	<b>Variáveis preditoras</b>	<b>Coeficiente <math>\beta</math></b>	<b>Erro-padrão</b>
		Porcentagem de pessoas vulneráveis a pobreza	0,04	0,01749

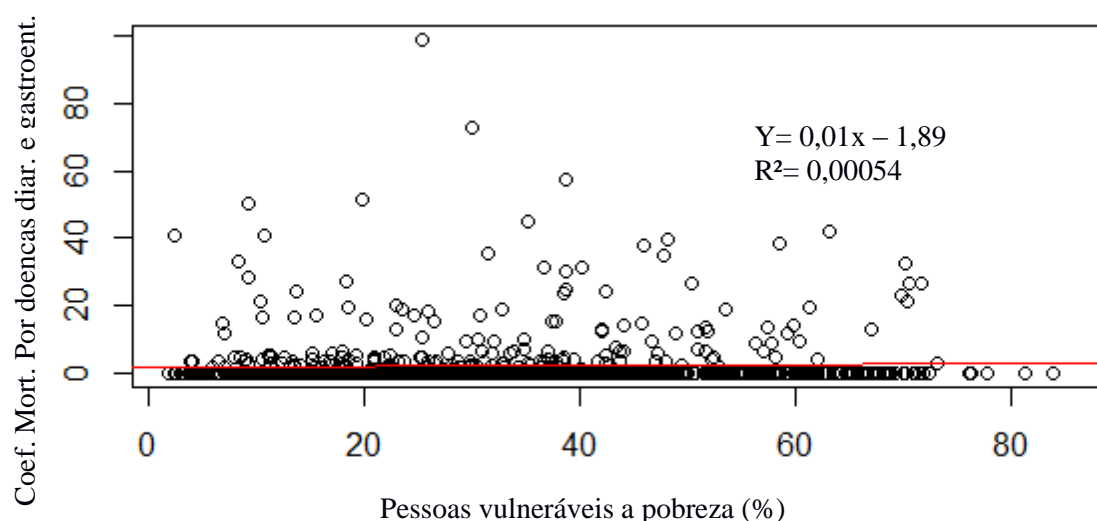
Fonte: Elaborado pelo autor, 2016. Nível de significância < 0,05.

De acordo com a Tabela 3, no “Ano” de 2010 observa-se um aumento de 0,18 casos por 100.000 habitantes no modelo 3 final.

Também, verifica-se a cada ponto percentual a mais de “Pessoas vulneráveis a pobreza” por microrregião, há um aumento de 0,04 casos por 100.000 habitantes no modelo 3 final.

A Figura 14 apresenta a variável preditora com maior significância para o desfecho “Coeficiente de mortalidade por doenças diarreicas e gastroenterite”, indicado pela análise de variância.

Figura 14 - Regressão linear simples entre “Coeficiente de mortalidade por doenças diarreicas e gastroenterite” e “Porcentagem de pessoas vulneráveis a pobreza”, COREDEs do RS (2000 e 2010)



Fonte: Elaborado pelo autor, 2016.

Observa-se na Figura 14, que a “Porcentagem de pessoas vulneráveis a pobreza” indica um acréscimo de 0,01 vezes no “Coeficiente de mortalidade por enfermidades diarreicas e gastroenterite”. O coeficiente de determinação não atingiu um valor satisfatório.

#### 5.1.4 Modelo para o “Coeficiente de morbidade por doenças diarreicas e gastroenterite” por 10.000 habitantes (modelo 4)

A técnica *stepwise backward selection* foi aplicada para o ajuste do modelo 4 final que foi composto pelas variáveis preditoras apresentadas na Tabela 4.

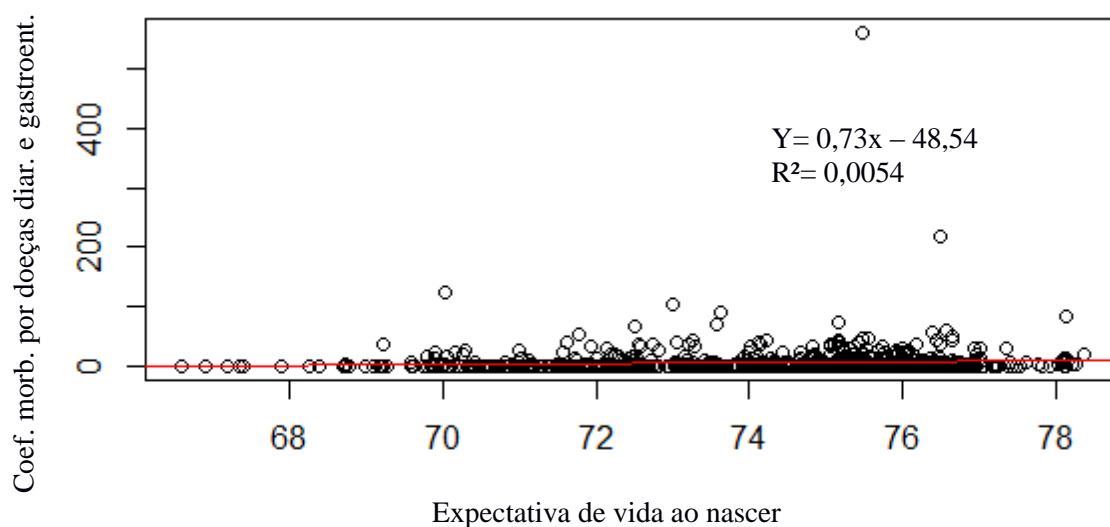
Tabela 4 - Regressão linear multivariada para o modelo 4 final dos COREDEs do Estado do RS nos anos 2000 e 2010

R <sup>2</sup>	Intercepto (coeficiente ± erro padrão)	Variáveis preditoras	Coeficiente $\beta$	Erro-padrão
0,0053	-48,29±23,75	Expectativa de vida ao nascer	0,72	0,32

Fonte: Elaborado pelo autor, 2016. Nível de significância < 0,05.

O modelo final foi composto somente por uma variável preditora significativa (“Expectativa de vida ao nascer”), indicando uma contribuição de 0,72 casos no modelo 4 final para as microrregiões do RS nos anos 2000 e 2010 (TABELA 4). A Figura 15 apresenta a reta ajustada do modelo.

Figura 15 - Regressão linear simples entre “Coeficiente de morbidade por doenças diarreicas e gastroenterite” e a “Expectativa de vida ao nascer”, COREDEs do RS (2000 e 2010)



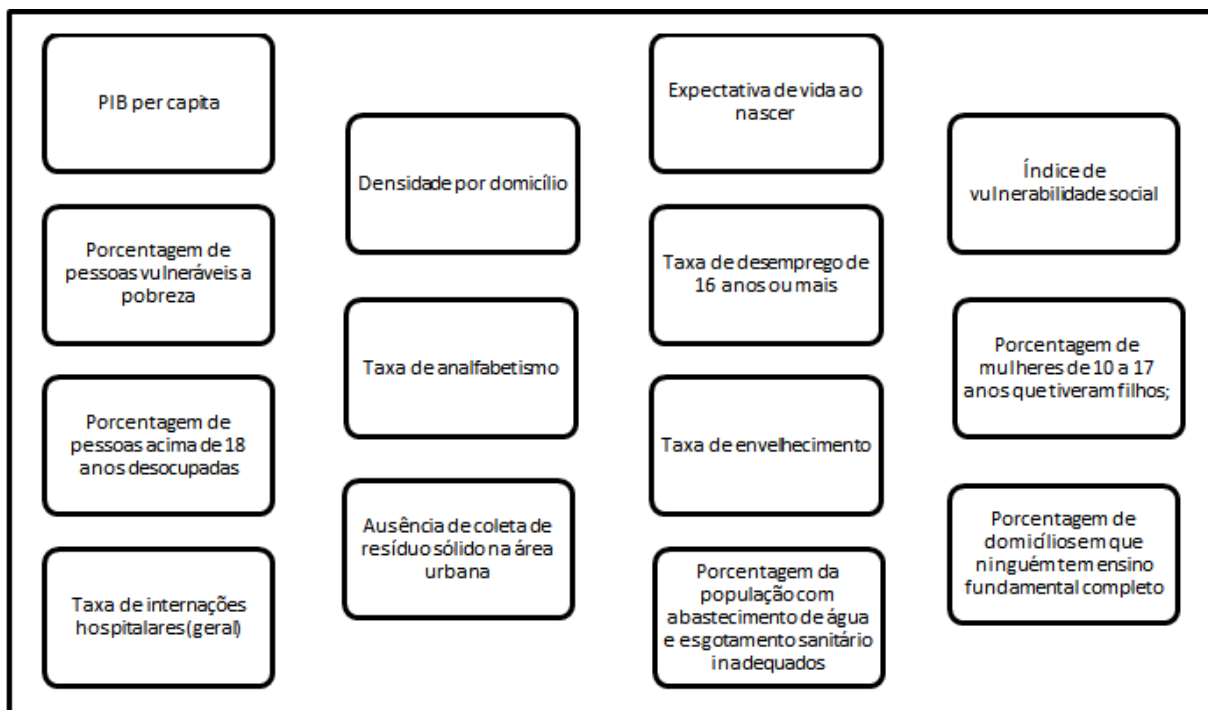
Fonte: Elaborado pelo autor, 2016.

Observa-se na Figura 15, que a “Expectativa de vida ao nascer”, indicou contribuição em 0,73 vezes no “Coeficiente de morbidade por doenças diarreicas e gastroenterite” para as microrregiões do RS nos anos 2000 e 2010. Porém, o coeficiente de determinação (0,54%) não foi satisfatório, indicando que essa variável preditora não descreve o desfecho unicamente.

## 5.2 MODELOS MULTINÍVEIS

As variáveis preditoras que possuem 20% ou mais de informação (*missing* < 20% do total de dados para a variável) estão apresentadas na Figura 16.

Figura 16 - Preditores com menos de 20% de dados faltantes para construção dos modelos multiníveis



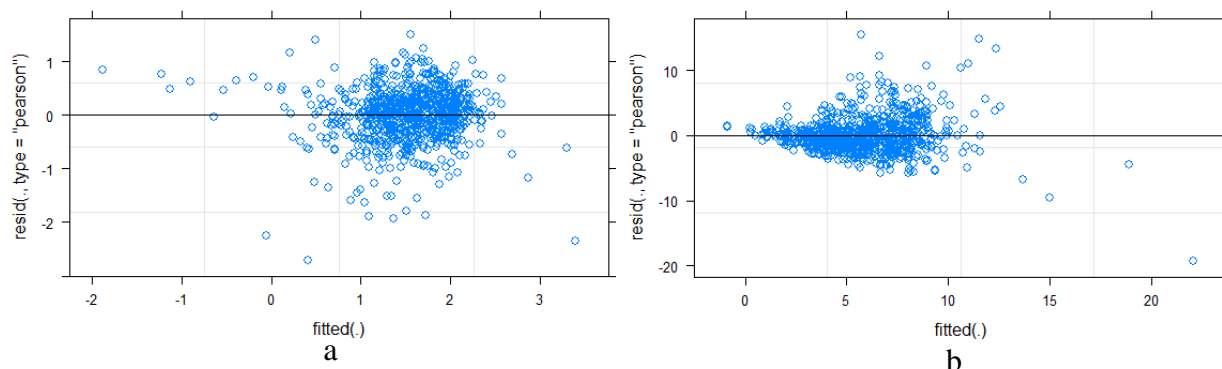
Fonte: Elaborado pelo autor, 2016.

### 5.2.1 Modelo para o desfecho “Coeficiente de morbidade por doenças infecciosas e parasitárias” por 1.000 habitantes (modelo 5)

A Figura 17 apresenta os gráficos de resíduos para os dois modelos propostos ao modelo 5 final considerando o desfecho em escala original e outro na logarítmica.



Figura 17 - Gráfico dos resíduos para modelo 5 com escala logarítmica (a) e original (b)



Fonte: Elaborado pelo autor, 2016.

A Figura 17 (a) apresenta o modelo com melhor ajuste dos pontos.

Os valores de AIC obtidos pelo algoritmo (descrito na metodologia) encontrados através de repetições, são apresentados na Tabela 5.

Tabela 5 - Valores de AIC para diferentes equações ajustadas no modelo 5

EQUAÇÃO AJUSTADA	AIC
morbdia1	1245,5
morbdia2	1245,2
morbdia3	1245,1
morbdia4	1244,9
morbdia5	1244,9
morbdia6	1244,9
morbdia7	1244,9
morbdia8	1244,9
morbdia9	1245,2
morbdia10	1244,9

Fonte: Elaborado pelo autor, 2016.

Observa-se na Tabela 5, que o menor valor de AIC em seis das dez repetições do algoritmo, apresentou-se com o valor de 1.244,9. A partir de então, verificou-se as estimativas das variáveis preditoras significativas ao modelo 5 final (Tabela 6).

Tabela 6 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 5 final dos COREDEs do Estado do RS nos anos 2000 e 2010

PARÂMETRO	Estimativa (Erro-padrão)
<b>FIXOS (Nível dos Municípios)</b>	
Intercepto	1,66 (0,04)
Ano	-0,25 (0,06)
Log (Porcentagem de crianças em domicílios em que ninguém tem fundamental completo)	-0,06 (0,03)
Log (Porcentagem de mulheres de 10 a 17 anos que tiveram filhos)	0,04 (0,02)
<b>FIXOS (Nível dos COREDEs)</b>	
Log (Taxa de Internações Hospitalares (geral))	0,46 (0,02)
<b>ALEATÓRIOS</b>	
	<b>Varição (DP)</b>
Região	0,02 (0,14)
Ano	0,03 (0,17)
Resíduos	0,26 (0,51)

Fonte: Elaborado pelo autor, 2016. DP = desvio-padrão.

Na Tabela 6, observa-se que a variável preditora “Ano” (variável binária), indica que de 2000 para 2010, houve uma redução de 0,25 casos no modelo 5 final. No caso do preditor “Porcentagem de crianças em domicílios em que ninguém tem fundamental completo”, em escala logarítmica, o aumento de um desvio-padrão corresponde a uma redução de 0,06 casos no modelo final analisado. Por outro lado, também em escala logarítmica, a cada desvio-padrão adicional na “Porcentagem de mulheres de 10 a 17 anos que tiveram filhos”, houve uma contribuição de 0,04 casos modelo 5 final. Por fim, o modelo ainda é composto pela “Taxa de internações hospitalares” (escala logarítmica) com base de dados ao nível dos COREDEs, o qual a cada aumento de um desvio-padrão há uma contribuição de 0,46 casos no modelo final estudado.

Sabe-se que em modelos de regressão, os valores observados serão próximos (não iguais) aos valores preditos (resultado da parte determinística do modelo). Esta distância entre o valor predito e o valor observado é denominada de “Resíduo”. Para verificar a distância da escala original da variável desfecho, pode-se utilizar a raiz quadrada da variância, a qual

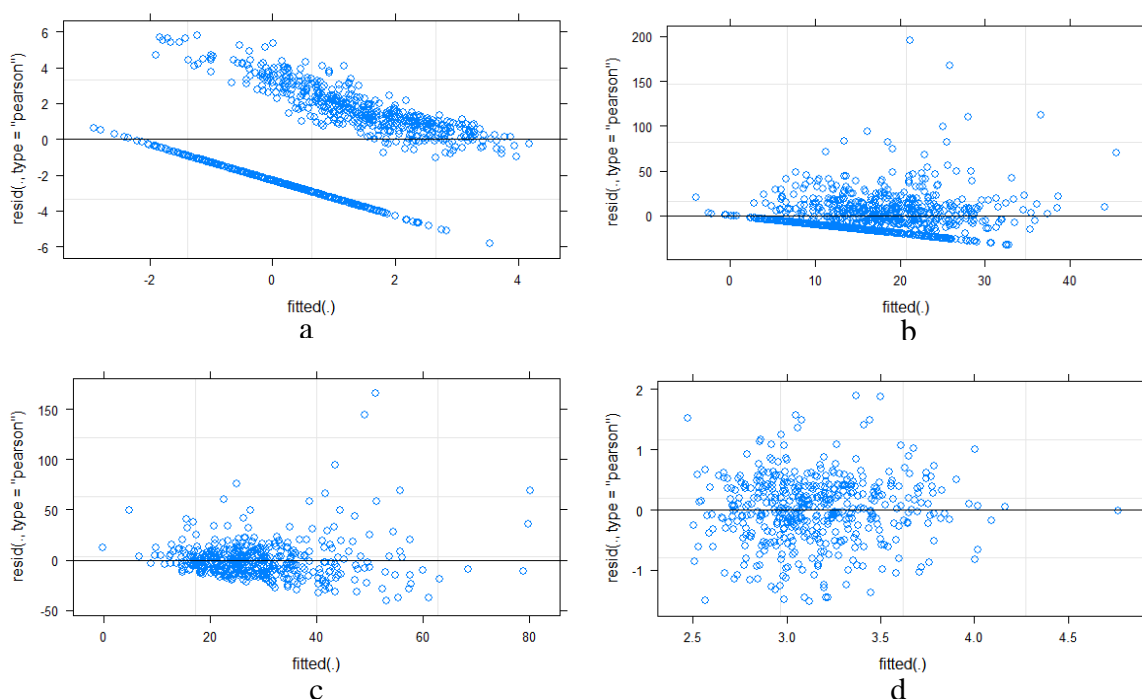
define o desvio-padrão. Com isto, observa-se na Tabela 6 que a distância média entre os valores preditos e os observados na escala original apresenta um desvio-padrão de 0,51.

Os modelos multiníveis permitem variar o intercepto em função de alguma variável agregadora (microrregiões). Os efeitos aleatórios por região são justamente o quanto que os interceptos de cada microrregião se distanciam da média geral. Como observado na Tabela 6, tem-se uma distância média (em desvios-padrão) de 0,14 (“Região”) e 0,17 (“Ano”) entre os valores preditos com os observados na escala original.

### 5.2.2 Modelo para o desfecho “Coeficiente de mortalidade por doenças infecciosas e parasitárias” por 100.000 habitantes (modelo 6)

A Figura 18 apresenta os resíduos para os quatro modelos ajustados ao modelo 6.

Figura 18 - Gráfico dos resíduos para o modelo 6 com função logarítmica e base de dados integral (a); função original e base de dados integral (b); função original e base de dados sem zeros (c); função logarítmica e base de dados sem zeros (d)



Fonte: Elaborado pelo autor, 2016.

O gráfico com melhor ajuste dos pontos está apresentado na Figura 18 (d), apresentando-se como uma função logarítmica e com a necessidade da exclusão de zeros na base de dados utilizada.

Na Tabela 7 são apresentados os valor de AIC para o modelo 6.

Tabela 7 - Valores de AIC para diferentes equações ajustadas no modelo 6  
(continua)

<b>DESFECHO SALVO COMO</b>	<b>AIC</b>
mortdiar1	870,0
mortdiar2	833,1
mortdiar3	830,6
mortdiar4	844,3
mortdiar5	833,7
mortdiar6	845,4
mortdiar7	846,8
mortdiar8	839,8
mortdiar9	849,9
mortdiar10	838,0
mortdiar11	837,2
mortdiar12	837,2
mortdiar13	841,8
mortdiar14	838,3
mortdiar15	849,8
mortdiar16	846,6
mortdiar17	846,6
mortdiar18	836,2
mortdiar19	838,5
mortdiar20	840,8
mortdiar21	836,6
mortdiar22	847,6
mortdiar23	843,4

<b>DESFECHO SALVO COMO</b>	<b>AIC</b>
mortdiar24	844,9
mortdiar25	836,6
mortdiar26	835,8
mortdiar27	837,8
mortdiar28	846,1
mortdiar29	842,6
mortdiar30	851,1
mortdiar31	838,9
mortdiar32	849,7
mortdiar33	838,9
mortdiar34	849,7
mortdiar35	839,3
mortdiar36	843,4
mortdiar37	841,4
mortdiar38	835,4
mortdiar39	835,8
mortdiar40	844,5
mortdiar41	830,9
mortdiar42	840,3
mortdiar43	852,4
mortdiar44	840,1
mortdiar45	845,8

Fonte: Elaborado pelo autor, 2016.

O menor valor de AIC em 45 repetições do algoritmo apresentou-se em 830,6. Assim, determina-se as estimativas para o modelo 6 final, em escore-z, apresentadas na Tabela 8.

Tabela 8 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 6 final para os COREDEs do Estado do RS, anos 2000 e 2010

<b>PARÂMETRO</b>	
<b>FIXOS (Nível dos Municípios)</b>	<b>Estimativa (Erro-padrão)</b>
Intercepto	3,11 (0,07)
Ano	0,07 (0,13)
Log (Porcentagem de pessoas vulneráveis a pobreza)	0,33 (0,07)
Porcentagem de pessoas acima de 18 anos desocupadas	0,15 (0,07)
Ausência de coleta de resíduo sólido na área urbana	0,05 (0,03)
Log (Porcentagem da população com banheiro e água encanada)	-0,09 (0,06)
Log (Porcentagem da população com abastecimento de água e esgotamento sanitário inadequados)	-0,15 (0,05)
IDHM	0,35 (0,12)
Densidade por Domicílio	0,18 (0,06)
Log (Expetativa de Vida ao Nascer)	0,09 (0,05)
IVS	0,11 (0,07)
Porcentagem de crianças em domicílios em que ninguém tem fundamental completo	0,08 (0,08)
Log (Porcentagem de mulheres de 10 a 17 anos que tiveram filhos)	-0,06 (0,03)
Log (Taxa de Desemprego)	-0,34 (0,06)
Taxa de envelhecimento	0,11 (0,04)
IDESE	0,12 (0,05)
<b>FIXOS (Nível dos COREDEs)</b>	<b>Estimativa (Erro-padrão)</b>
Log (Hospitais)	-0,09 (0,05)
Leitos	0,10 (0,04)
<b>ALEATÓRIOS</b>	<b>Variação (DP)</b>
Região	0,0001957 (0,01399)
Ano	0,0089889 (0,09481)
Resíduos	0,3148143 (0,56108)

Fonte: Elaborado pelo autor, 2016. DP = desvio-padrão.

As variáveis apresentadas na Tabela 8 foram padronizadas, por este motivo, o aumento ou diminuição corresponde a desvios-padrão. Com base nisto, as variáveis que permaneceram no modelo 6 final e estão associadas a um aumento do desfecho foram: “Porcentagem de pessoas vulneráveis a pobreza” (escala logarítmica) (aumento de 0,33 casos a cada 100.000 habitantes para cada ponto percentual adicional), “Porcentagem de pessoas acima de 18 anos desocupadas” (aumento de 0,15 casos a cada 100.000 habitantes), “Ausência de coleta de resíduos sólido na área urbana” (aumento de 0,05 casos a cada 100.000 habitantes), “IDHM” (aumento de 0,35 casos a cada 100.000 habitantes), “Densidade por domicílio” (aumento de 0,18 casos a cada 100.000 habitantes), “Expectativa de vida ao nascer” (escala logarítmica) (0,09 casos a cada 100.000 habitantes), “IVS” (0,11 casos a cada 100.000 habitantes), “Porcentagem de crianças em domicílios em que ninguém tem fundamental completo” (aumento de 0,08 casos a cada 100.000 habitantes), “Taxa de envelhecimento” (0,11 casos a cada 100.000 habitantes), “IDESE” (0,12 casos a cada 100.000 habitantes), “Leitos hospitalares” (0,10 casos a cada 100.000 habitantes) e por fim, a variável preditora “Ano” (variável binária), indica que a cada aumento de um desvio-padrão, houve um acréscimo de 0,07 casos do modelo 6 final.

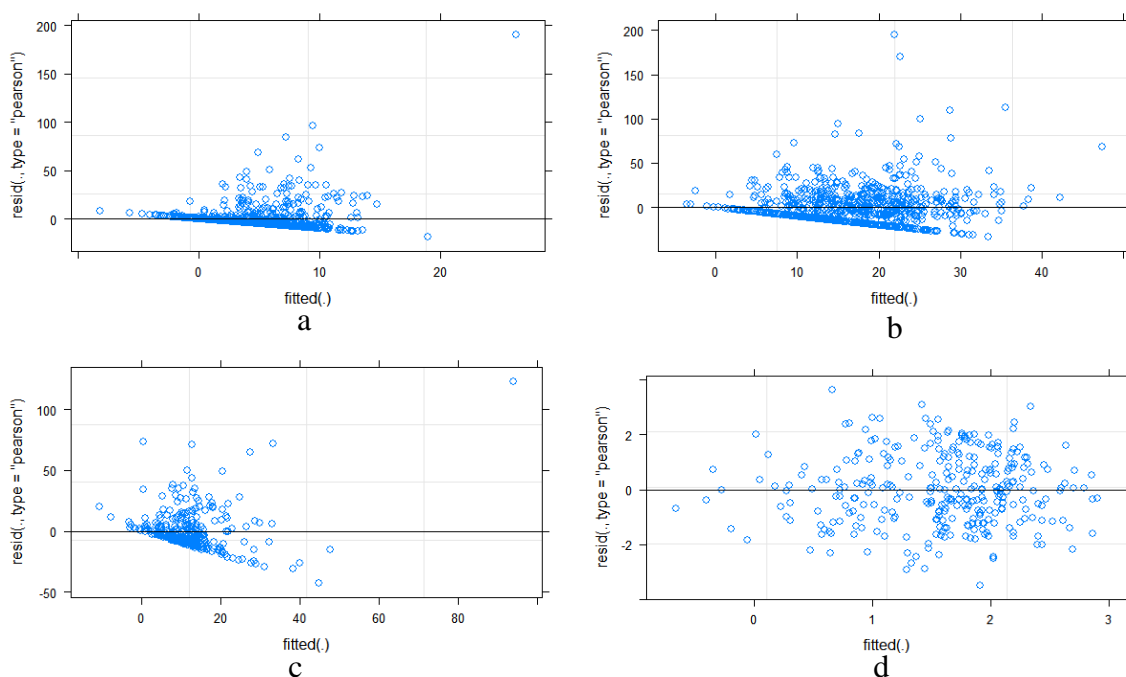
Por outro lado, as variáveis associadas à redução do modelo 6 final foram: “Porcentagem da população com banheiro e água encanada” (escala logarítmica) (redução de 0,09 casos a cada 100.000 habitantes), “Porcentagem da população com abastecimento de água e esgotamento sanitário inadequados” (escala logarítmica) (redução de 0,15 casos a cada 100.000 habitantes), “Porcentagem de mulheres de 10 a 17 anos que tiveram filhos” (escala logarítmica) (redução de 0,06 casos a cada 100.000 habitantes), “Taxa de desemprego” (escala logarítmica) (redução de 0,34 casos a cada 100.000 habitantes) e “Número de hospitais” (escala logarítmica) (redução de 0,09 casos a cada 100.000 habitantes).

Observa-se ainda na Tabela 6 que a distância média entre os valores preditos e os observados na escala original apresenta um desvio-padrão de 0,56. Também, o modelo apresenta uma distância média com desvios-padrão de 0,01 (Região) e 0,09 (Ano) entre os valores preditos com os observados na escala original.

### 5.2.3 Modelo para o desfecho “Coeficiente de morbidade por doenças diarreicas e gastroenterite” por 10.000 habitantes (modelo 7)

A Figura 19 apresenta os gráficos de resíduos para o modelo 7.

Figura 19 - Gráfico dos resíduos para modelo 7 com função original e base de dados integral (a); com função logarítmica e com base de dados integral (b); com função original e base de dados sem zeros (c); com função logarítmica e base de dados sem zeros (d)



Fonte: Elaborado pelo autor, 2016.

O modelo com melhor ajuste dos pontos é visualizado na Figura 19 (d), apresentando este uma função logarítmica e com a necessidade de exclusão de zeros em sua base de dados utilizada.

Na Tabela 9 são apresentados os valores de AIC para o modelo 7.

Tabela 9 - Valores de AIC para diferentes equações ajustadas no modelo 7 (continua)

EQUAÇÃO AJUSTADA	AIC
morbigestro1	1084,8



<b>EQUAÇÃO AJUSTADA</b>	<b>AIC</b>
morbigastro2	1084,6
morbigastro3	1084,4
morbigastro4	1084,6
morbigastro5	1087,8
morbigastro6	1091,8
morbigastro7	1084,9
morbigastro8	1084,1
morbigastro9	1087,5
morbigastro10	1088,1
morbigastro11	1087
morbigastro12	1084,1
morbigastro13	1084,1
morbigastro14	1085
morbigastro15	1086,3
morbigastro16	1084,1
morbigastro17	1083,9
morbigastro18	1087,5
morbigastro19	1086,8
morbigastro20	1084,1

Fonte: Elaborado pelo autor, 2016.

Observa-se na Tabela 9, que o menor valor de AIC em 20 repetições do algoritmo, apresentou-se em 1083,9. Assim, determinaram-se as estimativas para os coeficientes do modelo 7 final (TABELA 10).

Tabela 10 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 7 final para os COREDEs do Estado do RS, anos 2000 e 2010

<b>PARÂMETRO</b>	<b>Estimativa (Erro-padrão)</b>
<b>FIXOS (Nível dos Municípios)</b>	
Intercepto	2,25 (0,22)
Ano	-0,95 (0,29)
PIB per capita	0,15 (0,09)
Porcentagem de pessoas vulneráveis a pobreza	0,50 (0,17)
Porcentagem de pessoas acima de 18 anos desocupadas	-0,46 (0,13)
Log (IDHM)	0,35 (0,18)
Log (Densidade por Domicílio)	-0,22 (0,12)
<b>FIXOS (Nível dos COREDEs)</b>	
Hospitais	0,31 (0,13)
Log (Leitos hospitalares)	-0,43 (0,13)
<b>ALEATÓRIOS</b>	
<b>Variação (DP)</b>	
Região	0,0001521 (0,012335)
Ano	0,0000006577 (0,000811)
Resíduos	1,68 (1,30)

Fonte: Elaborado pelo autor, 2016. DP = desvio-padrão.

Na Tabela 10, verifica-se que ocorre uma redução de 0,95 casos por 10.000 habitantes para cada aumento de uma unidade no desvio-padrão da variável preditora “Ano”, comparando-se os anos de 2000 e 2010. Ainda, as variáveis que contribuíram de maneira positiva ao modelo: “PIB per capita” (contribuição de 0,15 casos), “Porcentagem de pessoas vulneráveis a pobreza” (contribuição de 0,50 casos), “IDHM” (escala logarítmica) (contribuição de 0,35 casos) e “Número de hospitais” com uma contribuição de 0,31 casos a cada aumento de desvio-padrão no modelo 7 final estudado.

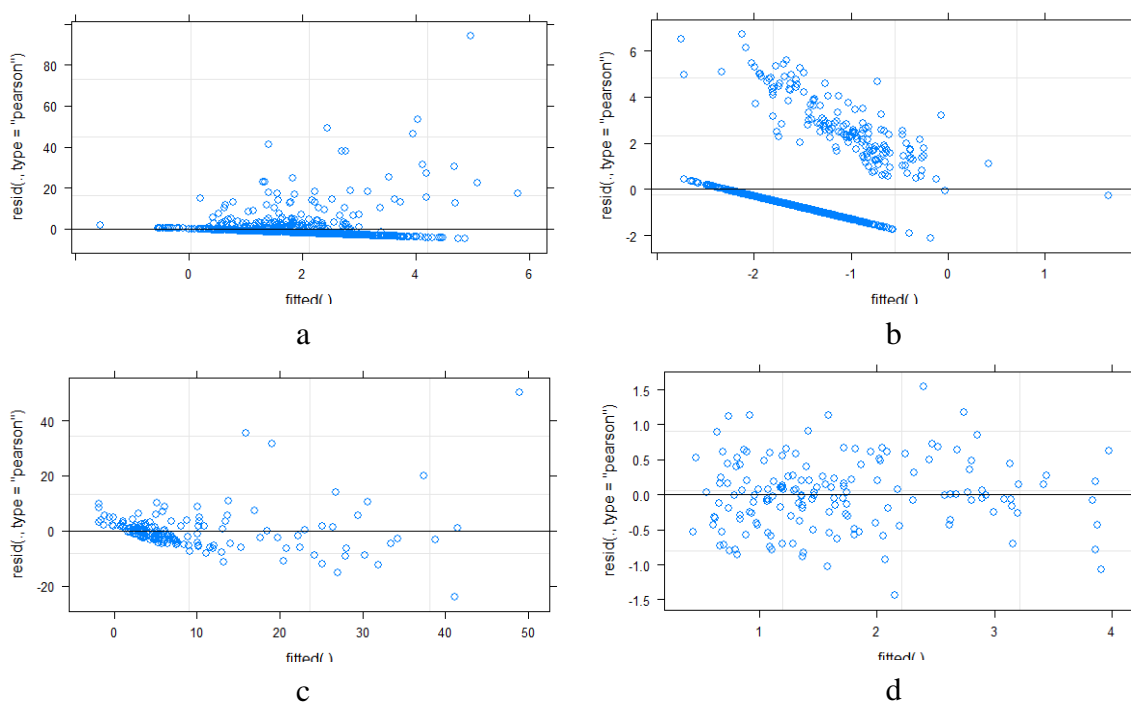
De maneira negativa (inversa) ocorreram as seguintes variáveis: “Porcentagem de pessoas acima de 18 anos desocupadas” (redução de 0,46 casos), “Densidade por domicílio” (escala logarítmica) (redução de 0,22 casos) e “Leitos hospitalares” (escala logarítmica) com uma redução de 0,43 casos no modelo 7 final.

Observa-se ainda na Tabela 10 que a distância média entre os valores preditos e os observados na escala original apresenta um desvio-padrão de 1,30. Também, o modelo apresenta uma distância média em desvios-padrão de 0,01 (Região) e 0,0008 (Ano) entre os valores preditos com os observados na escala original.

#### 5.2.4 Modelo para o desfecho “Coeficiente de mortalidade por doenças diarreicas e gastroenterite” por 100.000 habitantes (modelo 8)

A Figura 20 apresenta os gráficos de resíduos para o modelo 8.

Figura 20 - Gráfico dos resíduos para o modelo 8 com função original e base de dados integral (a); com função logarítmica e base de dados integral (b); com função original e base de dados sem zeros (c); com função logarítmica e base de dados sem zeros (d)



Fonte: Elaborado pelo autor, 2016.

O modelo com melhor ajuste dos pontos é apresentado na Figura 20 (d).

Os valores de AIC estão apresentados na Tabela 11.

Tabela 11 - Valores de AIC para diferentes equações ajustadas no modelo 8

<b>EQUAÇÃO AJUSTADA</b>	<b>AIC</b>
mortgastro1	302,5
mortgastro2	302,5
mortgastro3	302,5
mortgastro4	302,5
mortgastro5	302,5

Fonte: Elaborado pelo autor, 2016.

Observa-se na Tabela 11 que o valor de AIC nas 5 repetições do algoritmo, manteve-se em 302,5. Assim, determinaram-se as estimativas do modelo, em escore-z, apresentadas na Tabela 12 para o modelo 8 final.

Tabela 12 - Estimativas dos parâmetros fixos e aleatórios encontrados para o modelo 8 final para os COREDEs do Estado do RS, anos 2000 e 2010  
(continua)

<b>PARÂMETRO</b>	<b>Estimativa (erro padrão)</b>
<b>FIXOS (Nível dos Municípios)</b>	
Intercepto	1,83 (0,09)
Ano	-0,36 (0,17)
Porcentagem de pessoas vulneráveis a pobreza	0,47 (0,12)
Log (Porcentagem de pessoas acima de 18 anos desocupadas)	-0,54 (0,08)
Log (Porcentagem da população com banheiro e água encanada)	0,23 (0,09)
Log (Porcentagem da população com abastecimento de água e esgotamento sanitário inadequados)	-0,13 (0,08)
Log (Densidade por Domicílio)	-0,40 (0,11)
Log (IVS)	0,28 (0,11)
Log (Porcentagem de mulheres de 10 a 17-anos que tiveram filhos)	-0,13 (0,06)
Taxa de Analfabetismo	0,21 (0,07)
Taxa de envelhecimento	0,16 (0,08)

<b>PARÂMETRO</b>	<b>Estimativa (erro padrão)</b>
<b>FIXOS (Nível dos Municípios)</b>	
<b>FIXOS (Nível das COREDEs)</b>	
Número de hospitais	-0,17 (0,06)
<b>ALEATÓRIOS</b>	
	<b>Varição (DP)</b>
Região	0,01 (0,12)
Ano	0,05 (0,23)
Resíduos	0,27 (0,52)

Fonte: Elaborado pelo autor, 2016. DP = desvio-padrão.

Observa-se na Tabela 12 que a cada aumento do desvio-padrão, a variável preditora “Ano” indica uma redução de 0,36 casos por 100.000 habitantes no modelo 8 final, comparando-se 2000 e 2010. No caso do preditor “Porcentagem de pessoas vulneráveis a pobreza”, o aumento de um desvio-padrão corresponde a uma contribuição de 0,47 casos no modelo analisado.

Os indicadores “Porcentagem de pessoas acima de 18 anos desocupadas” (escala logarítmica), “Porcentagem da população com abastecimento de água e esgotamento sanitário inadequados” (escala logarítmica), “Densidade por domicílio” (escala logarítmica), “Porcentagem de mulheres de 10 a 17 anos que tiveram filhos” (escala logarítmica) e “Número de hospitais” contribuíram de maneira inversa a cada aumento no desvio padrão no modelo 8 final, sendo eles, respectivamente: 0,54; 0,13; 0,40; 0,13 e 0,17.

Ainda, verifica-se na Tabela 12 que preditores como “Porcentagem de pessoas vulneráveis a pobreza” (acrécimo de 0,47 casos), “Porcentagem da população com banheiro e água encanada” (escala logarítmica) (acrécimo de 0,23 casos), “IVS” (escala logarítmica) (acrécimo de 0,28 casos), “Taxa de analfabetismo” (acrécimo de 0,21 casos) e “Taxa de envelhecimento” (acrécimo de 0,16 casos) contribuíram de maneira positiva ao modelo analisado.

Dentre os principais resultados para os quatro modelos que descrevem os coeficientes de morbidade e mortalidade nas microrregiões do RS, tem-se que o modelo 5 final, foi composto por três preditores, sendo que um dos preditores com sinal inverso do esperado ao desfecho. O modelo 6 final, com 16 preditores, sendo que oito preditores resultaram no sinal

esperado ao desfecho. O modelo 7 final compôs-se de cinco preditores contrários ao desfecho esperado. Para o modelo 8 final, tem-se dez preditores, com seis preditores com sinal inverso do esperado ao desfecho.

Sugere-se como resposta a contrariedade dos sinais em alguns dos preditores dos modelos propostos, o fato de que as variáveis preditoras utilizadas no estudo possam ter baixa influência nas questões relacionadas aos desfechos modelados. Outro fator importante na elaboração dos modelos, é que não levou-se em consideração o peso dos municípios com seu respectivo número de habitantes, visto que, encontram-se grandes diferenças entre uma cidade pequena que não apresenta, por exemplo, tratamento de água, e mesmo assim há um nível significativo de qualidade de vida no local, quando se comparado a grandes centros, que com a mesma insuficiência no serviço, é fator determinante para uma epidemia local.

## **6 CONSIDERAÇÕES FINAIS**

Além da análise dos dados por modelos multiníveis, a contribuição do estudo é a disponibilização de um banco de indicadores de saúde e ambiente para as microrregiões do RS, os quais estarão disponíveis para outros estudos, tanto acadêmicos como para consulta da comunidade em geral.

Como trabalhos futuros sugere-se aplicar modelos multiníveis, considerando os mesmos desfechos deste estudo para crianças de até quatro anos de idade nas microrregiões do Estado do RS. Esta faixa etária é justificada por vários autores como aquela mais vulnerável aos índices de morbidade/mortalidade quando relacionada às questões de saneamento básico inadequado. Também, sugere-se o ajuste de modelos considerando outras variáveis preditoras no campo da saúde ambiental.

Pretende-se na sequência do estudo, considerar o peso dos municípios na elaboração dos modelos estatísticos, visto que, por exemplo, com a mesma insuficiência de determinado serviço de saneamento básico no local, há uma relativa diferença na qualidade de vida das pessoas em cidades pequenas, quando se comparado a grandes centros urbanos.

Por fim, percebeu-se a aplicabilidade da técnica de modelagem multinível para investigar e monitorar relações entre indicadores de saúde e ambiente.

## REFERÊNCIAS

ALLEBRANDT, S. L.; SIEDENBERG, D. R.; SAUSEN, J. O.; DECKERT, C. T.. Gestão Social e cidadania deliberativa: uma análise da experiência dos Coredes no Rio Grande Do Sul, 1990-2010. **Cadernos EBAPE.BR**, Rio de Janeiro, v.9,n.3, p. 914-945, Set.2011.

ATLAS DO DESENVOLVIMENTO HUMANO NO BRASIL 2013 (ATLAS BRASIL). Disponível em: <[www.atlasbrasil.org.br](http://www.atlasbrasil.org.br)>. Acesso em: 15 de março de 2016.

ANDREAZZI, M. A. R.; BARCELLOS, C.; HACON, S. Velhos indicadores para novos problemas: a relação entre saneamento e saúde. **Revista Panamericana de Salud Publica**, Washington, v. 22, n. 3, p. 211–17, 2007.

ATLAS SOCIOECONÔMICO DO RIO GRANDE DO SUL. Disponível em: <<http://www.scp.rs.gov.br/atlas>>. Acesso em: 15 de abril de 2016.

ATLAS DA VULNERABILIDADE SOCIAL. Disponível em: < <http://ivs.ipea.gov.br/ivs/>>. Acesso em: 05 de setembro de 2016.

BARROS, A. J. D. **Modelos multinível: primeiros passos**. Universidade de Pelotas: Departamento de Medicina Social. Faculdade de Medicina. Janeiro de 2001.

BATTISTI, I. D. E. **Correlação e regressão linear simples**. Cerro Largo, 2013. Disponível em: < [https://moodle-antigo.uffs.edu.br/file.php/4236/Apostila\\_Analise\\_de\\_Correlacao\\_e\\_Regressao.pdf](https://moodle-antigo.uffs.edu.br/file.php/4236/Apostila_Analise_de_Correlacao_e_Regressao.pdf)>. Acesso em 11 de dezembro de 2016.

BERTSIMAS, D.; TSITSIKLIS, J. Simulated Annealing. **Statistical Science**, v. 8, n 1, p. 10-15, 1993.

BRASIL. Ministério da Saúde. Instrução Normativa nº 01. **Regulamenta a Portaria GM/MS 1.172/2004 no que se refere às competências da União, estados, municípios e Distrito Federal na área de vigilância em Saúde Ambiental**. Diário Oficial da República Federativa do Brasil, Poder Executivo, Brasília, DF, 08 de março de 2005.

BRASIL. Ministério da Saúde. Fundação Nacional de Saúde. **Programa de Cooperação Técnica: estudos e pesquisas**. Brasília: Funasa, 2006. 40 p.



BRASIL. Ministério da Saúde. Fundação Nacional de Saúde. **Diretrizes de educação em saúde visando à promoção da saúde:** documento base – documento I. Brasília: Funasa, 2007. 70p.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância em Saúde Ambiental e Saúde do Trabalhador. **Saúde ambiental: guia básico para construção de indicadores.** Brasília: Ministério da Saúde, 2011. 124p.

CALIJURI, M. L. et al. Estudo de indicadores de saúde ambiental e de saneamento em cidade do Norte do Brasil. **Revista Engenharia Sanitária e Ambiental**, Rio de Janeiro, v. 14, n. 1, p. 19-28, 2009.

CASTRO, A. G.; DUARTE, A.; SANTOS, T. R. **O Ambiente e a Saúde.** Lisboa. Instituto Piaget, 2003. 435 p.

COSTA, S. S. da et al. Indicadores epidemiológicos aplicáveis a estudos sobre a associação entre saneamento e saúde de base municipal. **Revista Engenharia Sanitária e Ambiental**, Rio de Janeiro, v. 10, n. 2, p. 118- 127, 2005.

DEPARTAMENTO DE INFORMÁTICA DO SUS (DATASUS). Disponível em: <[www.datasus.gov.br](http://www.datasus.gov.br)>. Acesso em: 18 de Fevereiro de 2016.

EMILIANO, P. C. **Fundamentos e aplicações dos critérios de informação: Akaike e Bayesiano.** Dissertação (Pós-graduação em Estatística e Experimentação Agropecuária). Universidade Federal de Lavras, 2009.

FEEDADOS. Fundação de Economia e Estatística. Disponível em: <http://feedados.fee.tche.br>. Acesso em: 20 de junho de 2016.

FERRÃO, M. E. **Introdução aos modelos de regressão multinível em educação.** Campinas/SP: Komedi, 2003.

FERREIRA, A. **Disciplina de modelos lineares 2012-2.** Disponível em: <[http://www.iceb.ufop.br/deest/p3f11\\_d3p4rt4m3nt03st/arquivos/0.164970001409694296.pdf](http://www.iceb.ufop.br/deest/p3f11_d3p4rt4m3nt03st/arquivos/0.164970001409694296.pdf)> Acesso em: 20 junho de 2016.

FRANCO NETTO, G. et. al. Impactos socioambientais na situação de saúde da população brasileira: Estudo de indicadores relacionados ao saneamento ambiental inadequado. **Revista Tempus Actas em Saúde Coletiva**, Brasília, v. 4, n. 4, p. 53-71, 2009.

FREITAS, C. M. de; OLIVEIRA, S. G. de; SCHÜTZ, G. E.; FREITAS, M. B. **Abordagem ecossistêmica para o desenvolvimento de indicadores de sustentabilidade ambiental e de saúde - região do médio Paraíba, Rio de Janeiro**. Relatório Técnico-científico, Rio de Janeiro, 2007.

HEIMERDINGER, A. P. **Neoplasia de mama no Rio Grande do Sul: uma análise por modelos multiníveis**. Dissertação (Mestrado em Modelagem Matemática). Departamento de Física, Estatística e Matemática—Ijuí/RS: Unijí - Universidade Regional do Noroeste do Estado do Rio Grande do Sul, 2011.

IPEADATA. Instituto de Pesquisas Econômica Aplicada. Disponível em: <http://www.ipeadata.gov.br/>. Acesso em: 20 de julho de 2016.

GOLDSTEIN, H. **Multilevel statistical models**. 3 ed. Edward Arnold, 2003.

MEDRONHO, R. A. et al. **Epidemiologia**. São Paulo: Atheneu, 2 edição, 2007. 790 p.

MORAES, A. B. de. **Baixo peso de nascidos vivos no Rio Grande Do Sul, Brasil : uma análise estatística multinível**. Tese (Doutorado em Epidemiologia). Faculdade de Medicina-Porto Alegre/RS: Universidade Federal do Rio Grande do Sul - UFRGS, 2007.

OLIVEIRA, A. C. S. de. **Introdução ao R**. Universidade Federal de Mato Grosso, Mato Grosso, 2011.

OLIVEIRA, M. L. C.; FARIA, S. C. Indicadores de saúde ambiental na formulação e avaliação de políticas de desenvolvimento sustentável. **Revista Brasileira de Ciências Ambientais**, n. 11, p. 16-22, 2008.

RADICCHI, A. L. A.; LEMOS, A. F. **Saúde Ambiental**. Belo Horizonte: Nescon/UFGM, Coopmed, 2009. 76p.

RIBEIRO, H. Saúde Pública e Meio Ambiente: evolução do conhecimento e da prática, alguns aspectos éticos. *Rev. Saúde e Sociedade*, v. 13, n. 1, p. 70-80, 2004.

SAUSEN, M. G. **Indicadores de saúde ambiental na região das Missões/RS**. 2014. 83f. Trabalho de conclusão de curso (Engenharia Ambiental) – Universidade Federal da Fronteira Sul, curso de graduação em Engenharia Ambiental, Cerro Largo, 2014.

SILVA, D. R. R. da. **Inter-relação entre indicadores socioeconômicos, ambientais, epidemiológicos e as doenças diarreicas agudas em menores de cinco anos, no estado do Pará**. Dissertação (Mestrado em Modalidade Profissional em Saúde Pública), Escola Nacional de Saúde Pública Sérgio Arouca, Rio de Janeiro, 2011.

SILVA, B. F. da; DINIZ, J.; BORTOLUZZI, M. A. **Minicurso de estatística básica: introdução ao software R**. Universidade Federal de Santa Maria, Santa Maria, 2009.

SOBRAL, T. E. L.; BARRETO, G. **Estudo dos métodos de seleção de modelos matemáticos ARX utilizando critérios de informação**. 63ª Reunião Anual da SBPC. Goiânia/GO: SBPC, 2011. Disponível em: <<http://www.sbpcnet.org.br/livro/63ra/resumos/resumos/5271.htm>>. Acesso em: 26 de novembro de 2016.

TEIXEIRA, J. C.; GUILHERMINO, R. L. Análise da associação entre saneamento e saúde nos estados brasileiros, empregando dados secundários do banco de dados indicadores e dados básicos para a saúde 2003– IDB 2003. **Revista Engenharia Sanitária e Ambiental**, Rio de Janeiro, v. 11, n. 3, p. 277-282, jul./set. 2006.

TEIXEIRA, J. C. **Saúde ambiental**. Juiz de Fora: Universidade Federal de Juiz de Fora, 2012.

TEIXEIRA, J. C.; PUNGIRUM, M. E. M. de C.. Análise da associação entre saneamento e saúde nos países da América Latina e do Caribe, empregando dados secundários do banco de dados da Organização Pan-Americana d Saúde – OPAS. **Rev. Bras. Epidemiol.** 8(4): 365-76, 2005.

VERZANI, J. **Getting Started with RStudio**. Estados Unidos: O'Reilly, 2011.

ZANINI, R. R. **Modelos multiníveis aplicados ao estudo da mortalidade infantil no Rio Grande do Sul, Brasil, de 1994 a 2004**. 2007. 200f. Tese (Doutorado em Epidemiologia). - Universidade Federal do Rio Grande do Sul, Faculdade de Medicina, Porto Alegre, 2007.